

Sahibi / Owner

Dr. İbrahim USLAN

Genel Yayın Yönetmeni / General Publication Director

Dr. Bünyamin CİYLAN

Editör / Editor

Dr. Bünyamin CİYLAN

Alan Editörleri / Field Editors

Dr. Aslıhan TÜFEKÇİ

Dr. Necmi ALTIN

Dr. Nurettin TOPALOĞLU

Dr. Mehmet ATAK

Dr. Hacer KARACAN

Dr. M. Rahmi CANAL

Dr. Neslihan GÖKTÜRK

Dr. Selçuk Kürşat İŞLEYEN

Dr. Süleyman TOSUN

Dr. Tarek BEJAOUİ

Yayın Kurulu / Editorial Board

Dr. Dimitrios KARAMANIS, University of Western Greece, Yunanistan

Dr. Eleonora GUSEINOVİENE, Klaipeda University, Litvanya

Dr. Erdoğan DOĞDU, TOBB ETÜ Üniversitesi, Türkiye

Dr. Fawzi M. AL-NAİMA, University of Bagdat, Irak

Dr. Filippo NERI, University of Naples Federico II, İtalya

Dr. Güngör BAL, Gazi Üniversitesi, Türkiye

Dr. Hasan OĞUL, Başkent Üniversitesi, Türkiye

Dr. Hüseyin HIZIROĞLU, Kettering University, ABD

Dr. Kasim Mousa Alwan AL-AUBİDY, Philadelphia University, ABD

Dr. M. Ali AKCAYOL, Gazi Üniversitesi, Türkiye

Dr. Metin GÜRÜ, Gazi Üniversitesi, Türkiye

Dr. Mohammad N. A. HAWLADER, International Islamic University, Malezya

Dr. Recep DEMİRCİ, Gazi Üniversitesi, Türkiye

Dr. Robert BELOIU, Pitești University, Romanya

Dr. Saqib JAVED, Chalmers University of Technology, İsveç

Dr. Timothy TEO, The University of Auckland, Yeni Zellanda

Dr. Youcef SOUFI, University of Tébessa, Cezayir

Teknik Koordinatör / Technical Coordinator

Arş. Gör. M. Hanefi CALP, Gazi Üniversitesi, Türkiye

Bilişim Teknolojileri Dergisi hakemli bir dergidir.

Journal of Information Technologies is a refereed journal.

Bilişim Teknolojileri Dergisi 3 ayda bir yayınlanmakta olup EBSCO, ULAKBİM ve Google Scholar tarafından taranmaktadır.

Journal of Information Technologies is published every 3 months and indexed by EBSCO, ULAKBİM and Google Scholar.

Yazışma Adresi / Contact Address

Tunus Cad. No: 35 Kavaklıdere Çankaya / ANKARA

Telefon/Telephone: 0312 202 38 01 - Faks/Fax: 0312 212 79 29 - e-posta/e-mail : btd@gazi.edu.tr

Çevrimiçi Hakem Değerlendirme Sistemi / Online Referee Evaluation System: <http://btd.gazi.edu.tr>

ISSN:1307-9697 / e-ISSN: 2147-0715

NİSAN 2017 / APRIL 2017

Bilişim Teknolojileri Dergisi

Amaç ve Kapsam

Bilişim Teknolojileri Dergisi orijinal ve yenilikçi araştırma sonuçlarını yayınlamaktadır. Bu dergi, bilişimin farklı alanlarındaki uzman ve akademisyenler için bir ortamdır. Buna ilaveten, bilgisayar ile ilgili konuların geniş uygulamalarına da bu anlamda açığız. Bu dergiye gönderilecek makaleler, aşağıdaki birkaç alt alanların (en azından) birinden olabilir.

- Adli Bilişim
- Bilgisayar Bilimleri
- Bilişim Sistemleri
- Sağlık Bilişimi
- Yönetim Bilişim Sistemleri
- Siber Güvenlik

Bilişim Teknolojileri Dergisi'nin her geçen gün kendini yenileyen bilişim teknolojileri alanındaki akademisyenler ve profesyonellere temel bir kaynak niteliğinde olması amaçlanmıştır. Dergiye gönderilecek makaleler; **Türkçe veya İngilizce** kabul edilmektedir. Eğer, gönderilecek makale Türkçe ise; başlık, özet ve anahtar kelimelerin İngilizcesi mutlaka olmalıdır.

Dergiye gönderilen materyaller; başka bir yayın, konferans bildirimleri vb. gibi yerlerde daha önce yayınlanmamış veya kabul edilmemiş orijinal araştırmaları içermelidir.

Bu derginin değerlendirme politikasına göre, gönderilen her bir makale, kurulda verilen ilgili alt alan editör tarafından en az üç bağımsız hakeme gönderilmektedir. Bu bir hakem değerlendirme sürecidir; bu yüzden, **unvan ve yazarların isimleri** kağıdın değerlendirme sürümüne **dahil edilmemelidir**.

Dergi, makaleleri sadece online sistemi yoluyla <http://dergipark.gov.tr/gazibtd> sitesinden kabul etmektedir. E-mail ile gönderilen makaleler kabul edilmemektedir.

International Journal of Informatics Technologies

Aim and Scope

Journal of Informatics Technology (JIT) publishes novel and innovative research results. This journal is a multidisciplinary forum for experts and academicians from different fields of informatics. In addition, wide applications of computer-related topics are also welcome in that sense. The papers submitted to this journal can be in (at least) one of the several subfields below:

- Computer Forensic
- Computer Sciences
- Informatics Systems
- Health Informatics
- Management Informations Systems
- Cyber Security

This journal aims to be a good forum for the academicians and vocational professionals and wishes to be one of the basic sources of knowledge on this area. All papers submitted to this journal must be written in Turkish or English with a English title, abstract and Keywords. The submitted materials to this journal must contain original researches, which was not published before or has not been considered in another publication, conference proceedings, etc...

According to the review policy of this journal, each submitted paper would be send to at least three independent referees by the associate editor of the relevant subfield, which is given in the committes. This is a **peer-review** process, thus **the titles and names** of the authors **would not be included** in the review version of the paper.

The journal only receives the papers via its specific online system from <http://dergipark.gov.tr/gazibtd>. No e-mail submission is accepted.

İÇİNDEKİLER / CONTENTS

Using Biometric Features on Long Distance Videos for Accurate Pedestrian Age Classification <i>Ömer Faruk İNCE, İbrahim Furkan İNCE, Jang Sik PARK</i>	123
Klasik ve Sezgisel Bulanık İkili Karşılaştırma ile Yazılım Geliştirme Projelerinin Maliyet Tahmini: Uygulama Örneği <i>Ayfer BAŞAR</i>	129
TinyOS Tabanlı Telsiz Duyarga Ağları için Bir Konumlandırma ve k -Bağıllık Denetleme Sistemi <i>Orhan DAĞDEVİREN, Vahid Khalilpour AKRAM</i>	139
Bulut Tabanlı Mobil Diyabet Kontrol Uygulaması: Mobil Diyabetim <i>Pelin YILDIRIM, Fatma BOZYİĞİT, Mehmet Hilal ÖZCANHAN, Semih UTKU</i>	153
Veri Analizinde İstatistik mi Veri Madenciliği mi? <i>İlkin Ecem EMRE, Çiğdem SELÇUKCAN EROL</i>	161
Scrum Metodu Kullanılarak Bir Mobil Uygulama Geliştirme Sürecinin Gerçekleştirilmesi <i>Mehmet ELİBOL, Çiğdem SELÇUKCAN EROL</i>	169
Büyük Veride Kişi Mahremiyetinin Korunması <i>Can EYÜPOĞLU, Muhammed Ali AYDIN, Ahmet SERTBAŞ, Abdül Halim ZAİM, Onur ÖNEŞ</i>	177
Vardiya Çizelgeleme Problemi ve Bir Örnek Uygulama <i>Emre VARLI, Tamer EREN</i>	185
İlişkisel ve İlişkisel Olmayan (NoSQL) Veri Tabanı Sistemleri Mimari Performansının Yönetim Bilişim Sistemleri Kapsamında İncelenmesi <i>Serdar ÖZTÜRK, Hatice Ediz ATMACA</i>	199
Attribute Reduction in Stochastic Information Systems Based On α -Dominance <i>Emel KIZILKAYA AYDOĞAN, Mihrimah ÖZMEN</i>	211
AOMDV Protokolünde Black Hole Ataklara Karşı Geliştirilmiş Güvenlik Uygulaması <i>Sinan TOKLU, Aziz AYDIN</i>	221
Bilgisayar Kullanıcılarına Yönelik Duygusal İfade Tespiti <i>Uğur AYVAZ, Hüseyin GÜRÜLER</i>	231

Uzun Mesafeli Hareketli Görüntülerde Yayaların Biyometrik Özelliklerine Göre Yaşlarının Sınıflandırılması

Ömer Faruk İNCE, İbrahim Furkan İNCE, Jang Sik PARK

Department of Electrical and Electronics Engineering, Kyungsoong University, Korea, Republic Of
farukince21@ks.ac.kr

(Geliş/Received:02.05.2016; Kabul/Accepted:17.04.2017)

DOI: 10.17671/gazibtd.309264

Özet— Video tabanlı insan tespiti, günümüzde bir hayli yaygın olan çalışmalardan biridir ve bu konu hakkında birçok çalışmalar ve tasarılar mevcuttur. Daha da ayrıntılı bir çalışma için; nesne tespit edildikten sonra, nesnelere sınıflandırılabilir veya takip edilebilir. Yetişkin ve çocuk sınıflandırması, sosyal güvenliğin sağlanması için, özellikle de günümüzde artan vakaları göz önünde bulundurduğumuzda, pek yararlı olabilir. Yapılan çalışmanın gayesi, videolardaki görüntülerden insanları yetişkin ve çocuk olarak iki sınıfa ayırmaktır. Öncelikle insan tespiti için Haar sınıflandırıcı kullanılmıştır. Bir sonraki adımda ise, kafa ve tüm vücut uzunluğu kullanılarak biyometrik bir oran çıkarımı yapılmıştır. Bu orana göre de tespit edilen insanın yetişkin veya çocuk olduğu belirlenmiştir. Sonuçlarımız göstermektedir ki, yetişkin sınıflandırmasındaki doğruluk payı %74.7 ve çocuk sınıflandırmasındaki doğruluk payı ise %68.1'dir.

Anahtar Kelimeler— Biyometrik Özellikler, Yaş Sınıflandırması, Yaya Algılama

Using Biometric Features on Long Distance Videos for Accurate Pedestrian Age Classification

Abstract—Pedestrian detection is popular topic currently, and there are various studies on image and video based pedestrian detection. After detection, it can be more specified as classification, tracking, etc. Classification of child and adult is useful for the social security, because the number of social crimes is increasing. The goal of this study is classification of pedestrian in two categories as child and adult. In this study, Haar cascade classifiers are used. At first, the full body and head area of the pedestrian are detected. Next, biometric ratio between height and head of pedestrian is calculated. Then, age is classified by comparing the biometric ratio between height of head and whole body. Experimental results showed that proposed biometric feature has a significant discriminant value for distinguishing the child from the adult on long distance videos. The accuracy rates are 74.7% for adult, and 68.1% for child.

Keywords— Biometric Features, Age Classification, Pedestrian Detection.

1. INTRODUCTION

Object detection in visual platforms has been studied for a long time. It can be used for pedestrian detection, vehicle detection, gender classification, age classification, etc. The main purpose of object detection is pinpointing the location of target in the scene. One of the most significant stages of object detection is extracting features. For accurate real-time object detection, features have to be robust, differential, and easy to calculate [1]. Classification of child and adult is useful for the social security, because the number of social crimes is increasing. For this classification, several studies have been introduced. However, those studies are based on near distance [2]. For far distance classification, biometric information is very applicable. The ratio between the heights of head and body is a good information for age

classification, because the size of head becomes optimum before puberty.

We developed a new method for age classification with biometric information. We used the length ratio between head and full-body. The statistical information is used to decide the threshold for classification.

In this paper, general overview and previous studies on age classification is referred in Section II. The proposed new approach for age classification is explained in detail, in Section III. Following that, experimental results with considerations are shown in Section IV. Lastly, we end the paper with conclusion in Section V.

2. AGE CLASSIFICATION

2.1. Haar-like Features

Paula Viola and Michael Jones [2] proposed an effective object detection method which uses Haar-like feature based cascade classifiers.

Basically, in this machine learning work, cascade function is trained with data as positive and negative images. Then it can be used for detecting objects in digital images.

First of all, algorithm needs various data, so a lot of positive and negative images should be collected to train classifier. After that, useful features are needed to be extracted from those images. In this point, Haar features [3] involve the process. Haar features can be considered as convolutional kernels, and to obtain every single feature, we need to subtract sum of pixels under white rectangle from sum of under black rectangle.

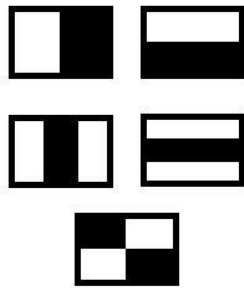


Figure 1. Haar-like features

To detect a good number of features, we have to use all possible sizes and locations of each kernel. That means, we will face a calculation issue which requires long time to calculate. Even though, we have 24×24 size window, it has more than 160000 number of features. For every single feature, sum of pixels are needed to be calculated. Also, for decreasing the time of calculation process, integral images were introduced. By using integral image, it doesn't matter how large the number of pixels are. It just simplifies the calculation, and calculates much faster.

In despite of everything, there is a big problem which is obtaining useful features. In other words, most of possible features are irrelevant. Here the answer is adaboost.

In general, boosting is a statistical method that uses the original distribution of positive and negative examples to compute simple rules also called weak classifiers and combines them to create a stronger classifier. Adaboost is most commonly used for binary classification, but with minor modifications it can also deal with multiple classes [4].

First of all, we apply all of those features, on training images. After that, algorithm finds useful features according to minimum error rate. Then, it can classify the object by using those features. Of course, there will be some errors or misclassifications. That is why we focus on selecting minimum error rates. Each image has equal probability at the beginning and after each steps are done, weight of misclassified images are increased. This process will be done iteratively until needed accuracy or error rate is obtained.

Final classifier is the last step of the algorithm. Sum of these weak classifiers are called final classifier. We call it weak, because it cannot classify images by itself. But when weak classifiers get together, they form strong classifier.

2.2. Overview to Previous Studies on Age Classification

Up to now, there are not so many studies published on the prediction of people's age from digital image or video. In the first instance, Kwon and Lobo [5] started predication of people's age from digital facial images and they described a method to classify people into three groups as babies, young adults, and senior adults. On a limited database, 100% accuracy is not applicable in real world applications. To complete similar age group classification, another technique which is based on craniofacial development theory was proposed [5]. But the approach is frustrated again due to limited data sets. With this way, it's very difficult to get robustness while we're dealing with real life conditions which has limitless circumstances.

There is also another method which is proposed by W.B. Horng, C.P. Lee and C.W. Chen [6]. This method put an extra group into account as middle aged adults to Kwon & Lobo's approach. The biggest restriction in this method is that they cannot detect children and younger adults who are under 40 years old. It can only detect babies. Despite everything, results were satisfactory and to find features, the analysis is pretty complex.

Lanitis [7] started up Active Appearance Models. To estimate the age, combined shape and texture parameters are dismissed by using classifiers. Simple quadratic fitting, shortest distance classifier, and Neural network are compared with age estimation accuracies of these classifiers. Geng [8] suggested static modeling on ageing patterns. It was based upon comparing input image with those patterns which consist of multiple images of various ages for each person.

3. PROPOSED VIDEO BASED CHILD AND ADULT CLASSIFICATION WITH BIOMETRY

Since studies on adult and child classification were introduced, different types of techniques are used. Some of them just preferred to use short distance detection

techniques, some of them used facial features to detect people according to their ages. In this study, our aim is different than the previous studies. We tried to classify people in long distance video scenes. For example, while one of previous studies was working on short distance classification (such as based on facial features) with webcams, this study focuses on detecting and classifying people on the street by using CCTV. Working on long distance can be considered as harder to implement, but it is definitely more useful for daily applications which are benefits of people.

Starting from this point of view, first of all, requirements for pedestrian detection from long distance scene should be determined. After that, classification requirements for child and adult can be discussed. Block diagram of the proposed algorithm can be shown as below:

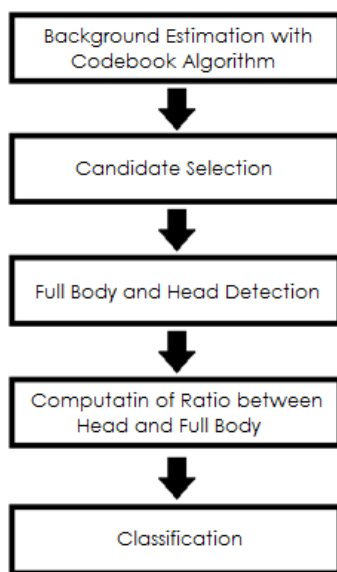


Figure 2. Block diagram of the proposed method

As it is seen from the block diagram, first, we need to separate foreground object from the scene by using codebook background subtraction. It is necessary to detect pedestrian faster and more accurate.

After first step, it is easy to detect pedestrian in the scene with using cascade classifiers based on adaboost algorithm. In the implementation part, more detail information, how cascade classifiers detect pedestrian, will be given.

Secondly, we use same way, which is using cascade classifiers to detect head of pedestrian. The only difference is that in the second step we were focusing to find full body in the whole scene. However in third step, we are trying to detect head of pedestrian inside full body.

After having full body information and head information for pedestrian, algorithm can be considered for classification. Here, the main idea of this study steps in.

There are medical facts that child and adult can be separated from each other by using biometry. In this study, the calculation of biometric ratio can be done by proportioning the length of full body to the length of head of the person, that is what we call computation of ratio. More details are given in the implementation part.

After we obtain threshold ratio by using biometry, we compare our ratio with threshold value, to see if pedestrian is child or adult.

3.1. Biometry for Classification

So far, we referred the proposition of previous algorithms for age classification and main framework of the proposed method. Now we are focusing on the last, but the most important step of the framework. That is classification of child and adult in digital videos.

In a digital video, there are various objects. Detecting human or human's head in the video, is applicable while ROI (Region of Interest) is human or humans' head. For implementation, we used Haar cascade classifier. After we detected the body, our new region of interest would be human body. After that, we needed to detect head of human. In this stage, we employed another Haar cascade classifier. By doing these steps, we finished first three steps which is shown in block diagram. For the third step, we needed to calculate biometry which is head and body ratio by equation:

$$\bar{O} = \lambda_H / \lambda_B \quad (1)$$

where λ_H and λ_B are height of head and body respectively, and \bar{O} means the ratio of biometry. There is a reason for using this ratio. It is medical fact that human's head size reaches optimum level before they become teenager. From this point of view, we can guess that this ratio can be discriminative between children and adults. We cannot measure people's heights from pixels in digital images. The reason is that it can be deceptive according to distance between pedestrian and camera. So, we need to solve this measurement problem. If we apply relative measurement which is related to proportioning head to body length by using classifiers, we would solve that problem. By this idea, we can classify that pedestrian is child or adult. If head over body ratio is bigger than 0.31, we can then classify that pedestrian is child. If that ratio is smaller than 0.31, that means pedestrian is adult. This is the main goal of the study. Also, with this proposition we can get low computation cost, and gain time. The reason is that we just focus on ratio, so we are not dependent to distance. Until now, the main idea of the proposed algorithm is explained, and threshold is mentioned. Now, we are focusing on how to calculate threshold value. To calculate threshold value λ_H / λ_B is measured on different videos which are taken in different mediums, with different positions.

Three main areas are assigned manually for observing relative measurement changes, and 4 different mediums are used to calculate optimum threshold value. Since it is calculated manually and four of all mediums have different conditions, the distance between bounds of areas might be slightly different than each other. Even though mediums, camera angles and distance are different, this situation gives better consideration points for obtaining optimum threshold value.

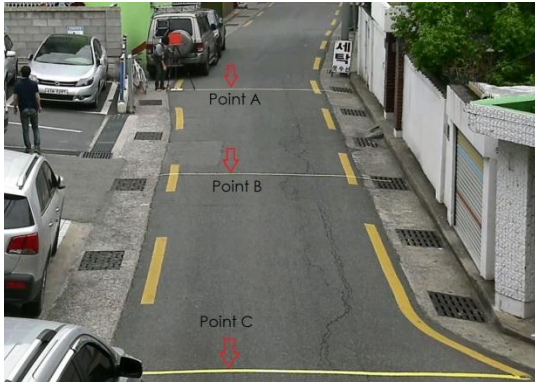


Figure 3. Separation points in the scene

These three areas are shown in figure, and the main idea is to see whether or not distance effect the ratio. For this reason, videos are collected and measurement is applied on both children and adults.



Figure 4. Separation points in the scene

The ratio λ_H / λ_B is calculated by counting pixel values manually. First, pixel values for height of head and then height of full body is calculated. All of values for children and adults are collected and their average values are calculated. The calculation rates for each groups in different mediums are given in the table.

Table 1. Ratios in different positions in the scene

Point	Medium 1			Medium 2		
	A	B	C	A	B	C
Child	0.30	0.32	0.32	0.32	0.33	0.32
Adult	0.24	0.23	0.23	0.27	0.26	0.26
Point	Medium 3			Medium 4		
	A	B	C	A	B	C
Child	0.31	0.31	0.33	0.32	0.32	0.30
Adult	0.28	0.27	0.28	0.28	0.27	0.29

According to table, average child and adult threshold values are 0.316 and 0.263 respectively. In the table, it is possible to see that adult's ratios are less stable than children. The main reason is that some of adults are really tall, and some of them are short. That issue affects the average ratio. However, the children's heights are quite

close to each other. Therefore, children's threshold value is taken as a reference threshold value. To be sure for exact threshold value, this calculation result is also verified by trial and error technique.

4. EXPERIMENTAL RESULTS

In this chapter, experimental results of this study will be focused on. First of all, four different videos are used to see how accurate this algorithm is. Also some of those videos' sizes are 1920×1080, and the rest of it is 1280×720. The aim was working on different media to get better idea for accuracy. Therefore, videos were taken in different conditions in Korea. Besides, each video has different view angle, illumination and distance from the target objects. OpenCV v2.4.9 via C++ on Microsoft Visual Studio 2010 is used for testing.

To classify pedestrians, we used two different Haar cascade classifiers which are for full body detection and head detection. 800 positive and 1600 negative images are used for full body XML file. Also HS.xml file, that Modesto Castrillon-Santana has copyrights, is used for head detection [9].

In this study, objects which are full body and head, are represented with different color rectangles. For full body detection, green color rectangle is used. However, when it comes to head detection, things are getting complicated. According to proposed algorithm, first we detect full body, then we focus on detecting head, after that we apply relative biometric measurement. If the ratio is bigger than threshold, that means the object is child. So, head rectangle will be shown as blue rectangle. If the ratio is smaller than threshold, that means the object is adult. So, head rectangle will be shown as red rectangle.

Results are given in two main categories: results of detecting and classification child and adult, and consideration on performance and accuracy rates of tested videos.



Figure 5. Child adult classification in different mediums

Results of four different videos are shown above. Two of them are adult detection, and the other two are child detection.

The reason of working on different media, is to see the real performance rate of this study. The first video is taken in subway, the second one is taken from street and the last two videos were taken from same street which is different street than the second video's medium. All media are in Busan, South Korea. Also, detection continuity is shown in the Figure 6.

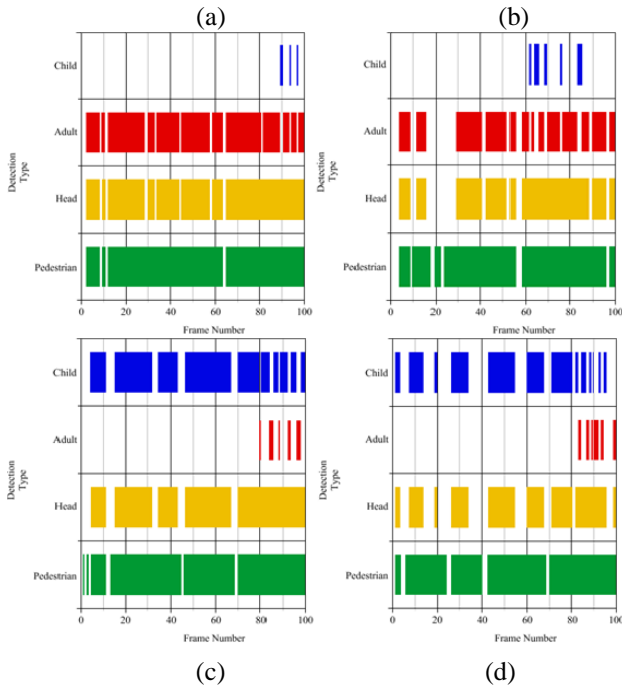


Figure 6. Detection continuity of experimented videos:
(a) Video 1, (b) Video 2, (c) Video 3, (d) Video 4

4.1. Considerations

Four different video results are shown in last section. According to the results, it is possible to say that classification for child and adult is working properly. However, numerical results are needed to see how accurate the proposed video based child and adult classification, is. Based on this, performance rates of each pedestrian, head, child and adult detection are given in the table within details, also average rates are calculated as well.

Table 2. Accuracy rate for detection and classification

	Full Body	Head	Adult	Child
Video 1 (Adult)	94.5%	90.8%	87.3%	2.7%
Video 2 (Adult)	89%	70.4%	62.1%	9.6%
Average (Adult)	91.75%	82.1%	74.7%	6.1%
Video 3 (Child)	93.7%	82.7%	6.6%	81.7%
Video 4 (Child)	92.5%	62%	8.6%	54.5%
Average (Child)	93.1%	72.35%	7.6%	68.1%

Detection and classification processes are implemented in narrow distance area where is between 2 meters and 10 meters away to camera. As the object goes away from camera, it is difficult to detect full body and head of human. That's why we work on narrower distance, and numbers of frames are not much.

Even though same algorithm is worked on those videos, the accuracy difference is relatively different. To answer that, there are couple reasons. Primarily, the biggest problem is having problem with detecting head. Because, once we find the head in the full body continuously, classification works successfully. However, failing for detecting head, decreases the total accuracy. For example, in the first video, woman is walking in subway station and since there is no other moving object, and her clothes' colors are totally different than background, the medium is totally proper to make pedestrian detection. Also, she is pretty close to the camera. For that reason, continuous head detection is possible and results of first video support this theory.

Same as the first two videos, the problem is same for the last two videos. That is difficulty of having continuous head detection. The XML file that we got to use on head detection, is trained for frontal angles. For theory, it can be useful, but in practice pedestrians always change the direction of their heads which causes decrements in the accuracy rate. Otherwise, this study's accuracy rate would be much better. Making a new proper XML file which is trained for every kind of head position, may solve this problem permanently.

5. CONCLUSION

Object detection is applicable for certain objects such as pedestrians, vehicles, or any other objects. Up to now, there are so many approaches and applications in various areas of computer vision about object detection.

In the process of time, children appear in daily life more than they have been. For example, children are under danger more than adults. Also criminal results show that child abduction rates are getting bigger and bigger every day. We can increase examples for sure. So we aimed to create an algorithm for child and adult classification for the benefits of children.

In this study, we proposed a new classification method to classify adult and child in videos with biometry. The absolute size of object is difficult to measure, since its size changes as the distance of its position changes in the scene. So, we used the information of head and full-body ratio instead of the exact size. Even though it cannot estimate the heights of pedestrian exactly, it showed that this approach is useful for adult and child classification. We tested four different videos which are taken in unequal conditions.

Experimental results showed that the proposed algorithm can classify pedestrian with 74.7% for adult, and 68.1%

for child. According to the performance graphs, continuous head detection, without any loss, is required. The reason is that, if we lose head detection, we will automatically lose the ratio calculation since we don't have any head information. So, creating more successful Haar cascade classifier for head, could be useful for better accuracy. In the future study, we will focus on improving the accuracy and expanding the application area such as night time age classification.

ACKNOWLEDGMENT

This work was supported by the Leading Human Resource Training Program of Regional Neo Industry through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT and Future Planning (No. 2016H1D5A1910536) and Brain Busan 21.

REFERENCES

- [1] K. Lee, C. H. Lee, S. A. Kim, and Y. H. Kim, "Fast Object Detection Based On Color Histograms and Local Binary Patterns", presented at the TENCON 2012, IEEE Region 10 Conference, Cebu, November, 2012.
- [2] P. Viola, M. Jones, "Rapid Object Detection Using a Boosted Cascade of Simple Features", Computer Vision and Pattern Recognition, 2001.
- [3] Internet: OpenCV Dev Team, "*penCV Dev Team, and Pattern Recogn*", http://docs.opencv.org/3.0-beta/doc/py_tutorials/py_objdetect/py_face_detection/py_face_detection.html, 2014.
- [4] G. R. Rakate, S. R. Borhade, P. S. Jadhav, M. S. Shah., "Advanced Pedestrian Detection System Using Combination of Haar-like Features, Adaboost Algorithm and Edgelet-Shapelet", IEEE Conference on Computational Intelligence and Computing Research, Coimbatore, India, 2012.
- [5] Y. H. Kwon and N. da Vitoria Lobo, "Age classification from facial images", Computer Vision and Image Understanding, 1999.
- [6] W. B. Horng, C. P. Lee, and C.W. Chen, "Classification of Age Groups Based On Facial Features", Tamkang Journal of Science and Engineering, vol. 4, no. 3, pp. 183-191, 2001.
- [7] A. Lanitis, C. Draganova, and C. Christodoulou, "Comparing Different Classifiers For Automatic Age Estimation", IEEE Transactions on Systems, Man and Cybernetics, vol. 34, no. 1, pp. 621-628, 2004.
- [8] X. Geng, Z. Zhou, Y. Zhng, G. Li, and H. Dai., "Learning from facial aging patterns for automatic age estimation", inProc. ACM Multimedia 2006, pp. 307-316, 2006.
- [9] M. Castrillon-Santana, O. Deniz- Suarez L., Anton -Canalis, and J. Lorenzo - Navarro, "Face and Facial Feature Detection Evaluation", VISAPP 2008, 2008.

Klasik ve Sezgisel Bulanık İkili Karşılaştırma ile Yazılım Geliştirme Projelerinin Maliyet Tahmini: Uygulama Örneği

Ayfer BAŞAR

Endüstri Mühendisliği, İstanbul Teknik Üniversitesi, İstanbul, Türkiye
ayferbasar@gmail.com

(Geliş/Received:01.06.2016; Kabul/Accepted:03.02.2017)

DOI: 10.17671/gazibtd.309269

Özet— Kaynak ve zaman planlamanın başarılı bir şekilde yürütülebilmesi için, yazılım geliştirme faaliyetlerinde maliyet tahminleme çok önemlidir. Maliyeti düşük tahmin edilen projelerde; gereğinden az kaynak atanması, gerçekçi olmayan planlama vb. sorunlar yaşanabilmekteyken tersi durumda kaynaklar israf edilmekte ve müşteri memnuniyeti düşebilmektedir. Tahmin edilen maliyetin doğruluğu, yazılım geliştirme projelerinin doğru bir şekilde önceliklendirilmesine, kaynakların etkin bir şekilde kullanılmasına, değişiklik isteklerinin başarılı bir şekilde yönetilmesine katkı sağlar. Bu nedenle, maliyetin gerçekçi bir şekilde öngörülmesini sağlayacak uygun yöntemin seçimi, tüm proje paydaşları açısından büyük önem arz etmektedir. Bu çalışmada, öncelikle deneyimli uzmanların görüşünden yararlanarak yazılım geliştirme maliyet tahmini için etkin kriterler belirlenmiş ve Klasik İkili Karşılaştırma yöntemi ile kriterlerin önem seviyeleri elde edilmiştir. Yazılım sektöründeki yoğun rekabet ve belirsizlikler, maliyet tahminleme için ihtiyaç duyulan tam ve kesin bilgi akışını zorlaştırmaktadır. Özellikle önceden belirlenmiş ölçekler kullanılarak yapılan değerlendirmelerde, kriterlerin göreceli öneminin ölçekte yer alan değerler ile tam olarak karşılanamaması, bulanık sayılar ile karar vermenin önemini arttırmaktadır. Bu nedenle, uzman görüşü yardımıyla belirlenen kriterlerin önemini tespiti için çalışmada ayrıca Sezgisel (Hesitant) Bulanık İkili Karşılaştırma önerilmiştir. Yöntemlerin etkinliğinin anlaşılması için, Türkiye’de hizmet vermekte olan bir yazılım şirketinin tamamlanmış projelerinin maliyeti tahmin edilmiştir. Gerçekleşen maliyetlere kıyasla, Sezgisel Bulanık İkili Karşılaştırma yönteminin klasik yöntemden daha başarılı sonuçlar verdiği görülmüştür.

Anahtar Kelimeler— yazılım maliyet tahminleme, karar verme, Klasik İkili Karşılaştırma, Sezgisel Bulanık İkili Karşılaştırma, uygulama

Estimating the Cost of the Software Projects with Traditional and Hesitant Fuzzy Pairwise Comparison: An Application

Abstract— Estimating the cost of software is a fundamental process in terms of resource and time planning. While underestimation causes committing fewer resources than the project needs and setting an unrealistic schedule, overestimation induces waste of resources and low customer satisfaction. By the help of efficient estimation methods, software projects can be prioritized efficiently, resources can be used effectively, and change request can be managed appropriately. Thus, choosing the appropriate cost estimation method is crucial for all the project shareholders. This paper presents a Traditional Pairwise Comparison (TPC) by using a number of criteria selected with the help of expert judgments working in banking and information technology industry in Turkey. It is clear that there may be uncertainties while evaluating the relative importance of two factors which may not be solved by TPC. In order to cope with this uncertainty in judgement, Hesitant Fuzzy Pairwise Comparison (HFPC) is also presented to estimate cost of software projects. In order to analyze the efficiency of proposed approaches, they are applied to a software cost estimation problem for a Turkish company. It is seen that HFPC provides more efficient estimations due to low deviation between the real and estimated effort in comparison with TPC.

Keywords— software cost estimation, decision making, Traditional Pairwise Comparison, Hesitant Fuzzy Pairwise Comparison, application.

1. GİRİŞ (INTRODUCTION)

Teknolojik gelişmelerin etkisiyle, yazılım geliştirme projelerinin işletmelerin rekabet gücü üzerindeki etkisi gün geçtikçe artmaktadır. Bu nedenle kurumlar, yazılım projelerinin zaman ve maliyet açısından etkin yönetimine önem vermek zorundadır. Projenin tahmini geliştirme maliyeti; çizelgeleme, kaynak tahsisi, performans yönetimi, müşteri memnuniyeti vb. kriterler açısından tüm paydaşlar açısından kritik öneme sahiptir. Ayrıca tahmin edilen maliyetin doğruluk seviyesi, projelerin öncelik ve planlama bilgisi üzerinde doğrudan belirleyici özelliğe sahiptir. Yazılım geliştirme maliyeti; efor (adam.gün / ay / hafta), süre (gün, ay, hafta) ya da parasal olarak tahmin edilebilmektedir.

Bu çalışmada, yazılım geliştirme projelerinin maliyet tahmini için önerilen yöntemler sunulmuştur. Maliyetin doğru tahmin edilmesi için alanında uzman olan yöneticilerin görüşü ile etkin kriterler belirlendikten sonra, Klasik ve Sezgisel Bulanık İkili Karşılaştırma yöntemleri kullanılarak kriterlerin göreceli önem seviyeleri elde edilmiştir. Elde edilen sonuçlar, gerçek hayat uygulaması üzerinde kıyaslanmış ve Sezgisel Bulanık İkili Karşılaştırma yönteminin daha başarılı sonuçlar verdiği görülmüştür. Makalenin devamı şu şekilde özetlenmiştir: 2. Bölümde maliyet tahminleme yöntemleri ile ilgili literatürde yer alan çalışmalar sunulacak, 3. Bölümde önerilen Klasik ve Sezgisel Bulanık İkili Karşılaştırma yöntemleri detaylandırılacak, 4. Bölümde önerilen yöntemler bir yazılım şirket üzerinde uygulanacak, 5. Bölümde ise sonuçlar ve gelecek çalışmalar aktarılacaktır.

2. MALİYET TAHMİNLEME YÖNTEMLERİ LİTERATÜR TARAMASI (LITERATURE REVIEW ON COST ESTIMATION METHODS)

Yazılım projelerinin maliyet tahmini için kullanılan yöntemler, temel olarak algoritmik olan ve olmayan şeklinde iki kategoride incelenmektedir. Uzman görüşü, algoritmik olmayan yöntemlerin en yaygın örneğidir. Konuya ilgili deneyimli uzmanların görüşünden yararlanabilme fırsatı, uzman görüşünün en önemli avantajıdır. Ancak zaman zaman uzmanlar tutarsız ya da taraflı davranabilmekte, böylece tahmin edilen maliyetler gerçekçi olmayabilmektedir [1, 2]. Uzman görüşü ile ortaya çıkabilecek tutarsızlıkları önlemek için Delphi Tekniği [3, 4] ve İkili Karşılaştırma [5] da maliyet tahmini için literatürde yaygın kullanılan algoritmik olmayan yöntemler arasında yerini almıştır. Algoritmik olmayan yöntemlerin diğer örneği olan Parkinson yaklaşımında maliyet, mevcut kaynaklar ve teslimat zamanına dayalı olarak öngörülmektedir. Parkinson yönteminde basitçe, teslimat zamanı 12 ay olan bir proje için 5 kişinin çalışması durumunda, çalışmanın maliyeti 60 adam-ay olarak tahmin edilmektedir [6].

Ancak algoritmik olmayan yöntemler ile elde edilen sonuçların, analitik ya da deneysel/matematiksel

modellere dayalı olmaması ve objektiflikten uzaklığı algoritmik yöntemlerin önemini arttırmıştır. Algoritmik yöntemlerde; x_1, x_2, \dots, x_n maliyet kriterleri ve $f(x_1, x_2, \dots, x_n)$ önceden belirlenen maliyet kriterlerinin kullanıldığı fonksiyon yardımıyla yazılım maliyeti tahmin edilir. Dolayısıyla maliyet kriterleri ve kullanılan fonksiyonun uygunluğu, gerçekçi tahmin açısından büyük öneme sahiptir. En yaygın maliyet kriterleri; ürün, bilgisayar, proje ve personeldir. Öte yandan algoritmik yöntemlerde; doğrusal, çarpımsal ve kuvvet şeklinde fonksiyon kullanılabilir. x_1, x_2, \dots, x_n maliyet kriterleri ve a_1, a_2, \dots, a_n ilgili kriterlerin önem düzeyini göstermek koşuluyla, doğrusal modellerde maliyet $a_0 + \sum_{i=1}^n a_i x_i$ formülü ile tahmin edilir [7]. Çarpımsal modelde maliyet tahmini için, $a_0 \prod_{i=1}^n a_i^{x_i}$ formülü kullanılır [8]. Kuvvet fonksiyon modellerinde ise maliyet; $a \times S^b$ formülü ile tahminlenir. Bu formülde, a ve b maliyet kriterleri, S ise kod büyüklüğünü temsil eder. Kuvvet fonksiyonlarının bilinen ve en yaygın örneği Constructive Cost Model (COCOMO) tekniğidir. Bu yöntemde, kod büyüklüğü olarak "satır sayısı" ve efor olarak "adam-ay" bilgisi referans alınır [9]. COCOMO yönteminin en büyük zorluğu, yazılım projesinin teslimatından önce kod satır sayısının doğru ölçülmesidir. Bu nedenle COCOMO yönteminde çoğu zaman, uzman görüşü ile Proje Değerlendirme ve Gözden Geçirme (Project Evaluation and Review Technique - PERT) tekniği kullanılarak kod büyüklüğü tahmin edilir. Böylece, uzmanlar 3 şekilde görüş bildirir: S_l en düşük, S_n en yüksek ve S_m en olası kod büyüklüğünü temsil eder. Daha sonra, $S = (S_l + 4S_m + S_n) / 6$ formülü ile kod büyüklüğü tahmin edilir. COCOMO yöntemi, genel olarak üç gruba ayrılır: i) Basit COCOMO, ii) Orta COCOMO, iii) COCOMO II. Basit COCOMO yöntemi, maliyet tahmini için $a \times (\text{KLOC})^b$ formülü kullanılan en eski COCOMO yaklaşımıdır [9]. Bu formülde, a ve b değerleri kodun karmaşıklığını temsil ederken, KLOC (Kilos of Lines of Code) ise 1.000 satırlık kod anlamına gelmektedir. Basit COCOMO yöntemindeki a ve b değerleri, yazılımın karmaşıklığına göre değişir. Bu parametreler sırasıyla; karmaşıklık düşükse 2,4 ve 1,05; orta derece ise 3,0 ve 1,15; yüksek ise 3,6 ve 1,20 değerini alır. Açıkça görülmektedir ki, basit COCOMO yönteminde maliyet sadece kod büyüklüğüne bağlıdır ve tahmini etkileyen diğer kriterler göz ardı edilmektedir. Bu nedenle, ilgili diğer parametreleri de göz önünde bulunduracak şekilde Orta COCOMO yöntemi geliştirilmiştir. Orta COCOMO yönteminde b parametresinin değerleri Basit COCOMO ile aynı iken; a parametresi değişmekte; ayrıca ürün, bilgisayar, proje ve personel başlığı altında toplam 15 kriterin farklı seviyelerdeki ağırlık değerlerine göre tahmin formülünde kullanılmak üzere Efor Ayar Kriteri (Effort Adjustment Factor) hesaplanmaktadır. Basit ve Orta COCOMO yöntemlerinin en önemli eksiği; yazılım projesinin tek ve homojen bir ürün olduğunu varsaymak, böylece büyük sistemlerin farklı karakterde çok sayıda alt sistem içerebileceğini göz ardı etmektir. Bu nedenle COCOMO II yöntemi geliştirilmiştir [10]. Bu yöntemde, b parametresi yeni 5 faktöre (hazır bulunma, geliştirme

esnekliği, risk yönetimi, takım kaynaşması ve süreç olgunluğu) göre farklı değerler almaktadır.

Son dönemlerde müşteri gereksinimlerindeki belirsizliğinin etkisiyle, işletmeler açısından hızlı üretim önem kazanmış ve çevik (agile) yaklaşımlar yaygın hale gelmiştir. Geleneksel yöntemlerden farklı olarak sık aralıklarla müşteriye ürün teslimatına dayalı çevik yaklaşımda, yazılım maliyet tahminleme için biyolojik bir sinir hücrelerinden oluşan paralel bilgi işleme yapıları olan yapay sinir ağları kullanılmıştır [11, 12]. Ayrıca özellikle problem boyutunun büyük olduğu durumlarda, en iyi ya da en iyiye yakın çözümleri bulmak için arama uzayını hızlı bir şekilde araştıran farklı meta sezgisel yöntemler de maliyet tahminleme probleminin çözümü için önerilmiştir. Örneğin Harmoni Arama Algoritması [13], Tabu Arama [14], Genetik Algoritma [14, 15, 16], Yapay Arı Algoritması [15] konuyla ilgili literatürde yer alan örneklerdendir. Harmoni Arama Algoritması, bir orkestrada görev alan müzisyenlerin çaldığı notalar ile armonik açıdan en iyi melodinin elde edilmesine dayalıdır. Tabu Arama yöntemi, yerel en iyi noktaya takılmayı engellemek için mevcut çözümü iyileştirmeyen hareketlere izin vermekte ve aramanın tarihçesini tutan tabu listelerinin kullanımı sayesinde önceki aşamalarda ziyaret edilen noktalara tekrar uğrama olasılığını düşürmektedir. Evrim teorisi esasına göre çalışan Genetik Algoritmada, potansiyel çözümler ikili olan ya da olmayan sistemlere dayalı veri yapısında basit diziler olarak şifrelenmekte ve söz konusu dizilere problemin çözümüne yönelik işlemler uygulanmaktadır. Yapay Arı Algoritması ise arı topluluklarının davranışlarını temel alan algoritmadır. Meta sezgisel yöntemlerin dışında, tek ve kesin değer yerine bir aralık kullanılarak bulanık mantık ile tahminleme yapılan uygulamalar da mevcuttur [17, 18].

3. İKİLİ KARŞILAŞTIRMA İLE KARAR VERME (DECISION MAKING WITH PAIRWISE COMPARISON)

Çok Kriterli Karar Verme (ÇKKV) yöntemleri; genellikle birden fazla sayıda ve çoğu zaman birbiri ile çelişen amaç ve kriterin olduğu, kriterler arasında tercih yapmanın zor olduğu durumlarda kullanılır. Böylece, ÇKKV problemlerinin sonuçları çoğunlukla karar vericilerin tercihlerine dayalıdır [19]. ÇKKV problemlerindeki kriterlerin önem düzeyini belirlemek için; uzman görüşünden yararlanılabilmesi, her defasında sadece iki kriterin karşılaştırılması, kolay ve anlaşılır olması vb. avantajları sayesinde, literatürde etkin bir yöntem olan Klasik İkili Karşılaştırma önerilmiştir [20]. Ancak, karar sürecindeki belirsizlik ve öznellik dolayısıyla Klasik İkili Karşılaştırma yönteminin etkin olmadığı durumlar söz konusu olabilmektedir. Bu nedenle literatürde, Bulanık İkili Karşılaştırma yöntemi önerilmiştir [21]. Bulanık İkili Karşılaştırma, tek ve kesin bir değer yerine bir aralığı kullanarak karar vericilere iyimser-orta-kötümser şeklinde farklı tercih sunma fırsatı sağlar. Ayrıca yakın zamandan

bu yana Sezgisel Bulanık İkili Karşılaştırma da literatürdeki yerini almıştır [22, 23].

2.1. Klasik İkili Karşılaştırma Yöntemi (Traditional Pairwise Comparison Method)

İkili Karşılaştırma, karar vericilerin yargısına göre her kriterin amaca katkısı açısından göreceli öneminin belirlendiği yöntemdir [24]. Öncelikle; uzman görüşünün toplanabilmesi için amaç, ana ve varsa alt kriterleri içeren karar hiyerarşisi hazırlanmaktadır. Toplam kriter sayısının n olduğu durumda, hiyerarşik yapıdaki kriterlerin göreceli ağırlıklarının belirlenmesi için, $n \times n$ boyutundaki karşılaştırma matrisi (A) yardımıyla karar verici tarafından ikili karşılaştırmalar yapılmaktadır. $i=1,2,3,\dots,n$ ve $j=1,2,3,\dots,n$ olmak üzere matris elemanları arasında $a_{ij}=1/a_{ji}$ ilişkisi bulunmakta olup $i=j$ olduğu için matrisin köşegenleri 1 değerini almaktadır.

$$A = \begin{bmatrix} 1 & a_{21} & a_{31} & \dots & a_{n1} \\ 1/a_{21} & 1 & a_{32} & \dots & a_{n2} \\ 1/a_{31} & 1/a_{32} & 1 & \dots & a_{n3} \\ \dots & \dots & \dots & \dots & \dots \\ 1/a_{n1} & 1/a_{n2} & 1/a_{n3} & \dots & 1 \end{bmatrix}_{n \times n} \quad (1)$$

Kriterlerin ikili karşılaştırması için [24] tarafından geliştirilen Tablo 1'deki önem ölçeği kullanılmaktadır.

Tablo 1: Önem Ölçeği [24] (Importance Scale)

Değer	Tanım
1	Elemanlar eşit öneme sahiptir.
3	1. eleman 2.'ye göre biraz daha önemli veya biraz daha tercih edilmektedir.
5	1. eleman 2.'ye göre fazla önemli veya fazla tercih edilmektedir.
7	1. eleman 2.'ye göre çok fazla önemli veya çok fazla tercih edilmektedir.
9	1. eleman 2.'ye göre aşırı derecede önemli veya aşırı derecede tercih edilmektedir.
2, 4, 6, 8	Ara değerler

İlgili matristeki her bir öğenin diğer öğelere göre önemini gösteren $n \times 1$ boyutundaki öz vektör, (2)'deki formüller aracılığıyla hesaplanır:

$$b_{ij} = a_{ij} / \left(\sum_{i=1}^n a_{ij} \right) \quad w_i = \left(\sum_{j=1}^n b_{ij} \right) / n \quad (2)$$

2.2. Bulanık İkili Karşılaştırma Yöntemi (Hesitant Pairwise Comparison Method)

ÇKKV problemlerindeki belirsizliği önlemek için bulanık küme teorisi önerilmiştir [25]. Bulanık küme teorisinde bir eleman, bir küme yerine k ($0 \leq k \leq 1$) derecesine ve "üyelik fonksiyonu" adındaki aralığa dahil olmaktadır (örn. üçgen üyelik fonksiyonu). Karar vericilerin belirsiz ortamlarda kriterler arasında tercih sunabilmesi için [25] tarafından bulanık ağırlıklandırma yöntemi önerilmiş ve

$a_{ij} = (l_{ij}, m_{ij}, u_{ij})$ şeklinde üçgen üyelik fonksiyonu kullanılmıştır. Bu gösterimde l alt limit, m orta değer ve u ise üst limiti temsil etmektedir. İlk olarak, karar vericiler Tablo 2’de yer alan bulanık ölçeğe uygun olarak tercihlerini belirtmekte, sonra bulanık karşılaştırma matrisi $A = (a_{ij})$ oluşturulmakta, en sonunda ise her kriter ve alternatif için bulanık önem seviyeleri tespit edilmektedir.

Tablo 2: Bulanık İkili Karşılaştırma Yönteminde Kullanılan Önem Ölçeği [26] (Importance Scale Used in Fuzzy Pairwise Comparison)

Önem	Dilsel Ölçek	Üyelik fonksiyonu	Aralık	Üçgen bulanık ölçek (l, m, u)
1	Eşit önem	$\mu_M(x) = (3-x) / (3-1)$	$1 \leq x \leq 3$	(1, 1, 1) (1, 1, 3)
3	Biraz daha önemli	$\mu_M(x) = (x-1) / (3-1)$ $\mu_M(x) = (5-x) / (5-3)$	$1 \leq x \leq 3$ $3 \leq x \leq 5$	(1, 3, 5)
5	Fazla önemli	$\mu_M(x) = (x-3) / (5-3)$ $\mu_M(x) = (7-x) / (7-5)$	$3 \leq x \leq 5$ $5 \leq x \leq 7$	(3, 5, 7)
7	Çok fazla önemli	$\mu_M(x) = (x-5) / (7-5)$	$5 \leq x \leq 7$	(5, 7, 9)
9	Aşırı önemli	$\mu_M(x) = (9-x) / (9-7)$ $\mu_M(x) = (x-7) / (9-7)$	$7 \leq x \leq 9$	(7, 9, 9)
Yukarıdaki önem seviyesi, i kriterinin j kriteri ile karşılaştırılması durumunda kullanılır. Aksi durumda, j kriteri tersine değeri alır.				Tersine değer $1/u, 1/m, 1/l$

2.2.1. Sezgisel Bulanık Kümeler (Hesitant Fuzzy Sets)

Karar vericiler tarafından kriterlerin göreceli öneminin belirlenmesi sırasında yaşanabilecek kararsızlık, bir elemanın üyelik derecesinin belirlenmesini zorlaştırmaktadır. Bu nedenle, [27] tarafından sezgisel bulanık kümeler tarif edilmiştir. Sezgisel bulanık kümelerin en önemli özellikleri şu şekildedir:

Tanım 1: Eğer X sabit bir küme ise, X 'in bağlı olduğu sezgisel bulanık küme $E = \{ \langle x, h_E(x) \rangle \mid x \in X \}$ ile gösterilir. Burada $h_E(x)$, E kümesindeki $x \in X$ elemanının $[0, 1]$ değerini alan olası üyelik derecesini temsil eder.

Tanım 2: Üst ve alt limitler, sırayla $h^-(x) = \min h(x)$ ve $h^+(x) = \max h(x)$ ile tespit edilir.

Tanım 3: h kümesinin tümleyeni, $h^c = U_{\in h} \{1-\gamma\}$ ile elde edilir. **Tanım 4:** Her biri sezgisel bulanık küme olan h, h_1 ve h_2 ile yapılan temel matematiksel işlemler şu şekildedir:

$$h^2 = U_{\gamma \in h} \{\gamma^2\} \quad (3)$$

$$\lambda h = U_{\gamma \in h} \{1 - (1-\gamma)^\lambda\} \quad (4)$$

$$h_1 \cup h_2 = U_{\gamma_1 \in h_1, \gamma_2 \in h_2, \max\{\gamma_1, \gamma_2\}} \quad (5)$$

$$h_1 \cap h_2 = U_{\gamma_1 \in h_1, \gamma_2 \in h_2, \min\{\gamma_1, \gamma_2\}} \quad (6)$$

$$h_1 \oplus h_2 = U_{\gamma_1 \in h_1, \gamma_2 \in h_2, \{\gamma_1 + \gamma_2 - \gamma_1 \gamma_2\}} \quad (7)$$

$$h_1 \otimes h_2 = U_{\gamma_1 \in h_1, \gamma_2 \in h_2, \{\gamma_1 \gamma_2\}} \quad (8)$$

Tanım 6: $(a_1, a_2, \dots, a_n) = \sum_{j=1}^n w_j b_j$ formülü aracılığıyla Ordered Weighting Averaging (OWA) operatörü hesaplanır. Bu formülde b_j ; a_1, a_2, \dots, a_n arasındaki j . en büyük değeri ve $[0, 1]$ arasındaki w_i ise i kriterinin önem düzeyini gösterir ($\sum_{j=1}^n w_j = 1$).

Tanım 7: Sezgisel bulanık kümelere bağlı olarak, karşılaştırmalı dilsel ifadeler $\tilde{A} = (a, b, c)$ şeklinde üçgen bulanık üyelik fonksiyon ile temsil edilir. $\tilde{A} = (a, b, c)$, şu şekilde hesaplanır:

$$a = \min \{ a_L^i, a_M^i, a_M^{i+1}, \dots, a_M^j, a_R^j \} = a_L^i \quad (9)$$

$$b = \text{OWA}_w (a_M^i, a_M^{i+1}, \dots, a_M^j) \quad (10)$$

$$c = \max \{ a_L^i, a_M^i, a_M^{i+1}, \dots, a_M^j, a_R^j \} = a_R^j \quad (11)$$

2.2.2. Sezgisel Bulanık İkili Karşılaştırma Yöntemi (Hesitant Fuzzy Pairwise Comparison Method)

Sezgisel Bulanık İkili Karşılaştırma yönteminin adımları şu şekildedir:

Adım 1: Tüm kriter ve alt kriterler için dilsel ölçeklere uygun olarak uzman görüşü toplanır, ikili karşılaştırma matrisleri oluşturulur.

Adım 2: Tablo 3’te yer alan ölçek kullanılarak, dilsel ifadeler üçgen bulanık sayılara dönüştürülür, uzman görüşü yardımıyla ikili karşılaştırma matrisi \tilde{A}^k oluşturulur ve farklı uzmanların görüşlerinin birbiri ile tutarlı olup olmadığı kontrol edilir.

Tablo 3: Sezgisel Bulanık İkili Karşılaştırma Yönteminde Kullanılan Ölçek (Scale Used in Hesitant Fuzzy Pairwise Comparison)

Dilsel İfade	Sembol	Üçgen Bulanık Sayı
Kesinlikle Çok Yüksek	KÇY	7,9,9
Çok Yüksek	ÇY	5,7,9
Aslında Yüksek	AY	3,5,7
Zayıf Derecede Yüksek	ZDY	1,3,5
Eşit Derecede Yüksek	EDY	1,1,3
Tam Eşit	TE	1,1,1
Eşit Derecede Düşük	EDD	1/3,1,1
Zayıf Derecede Düşük	ZDD	1/5,1/3,1
Aslında Düşük	AD	1/7,1/5,1/3
Çok Düşük	ÇD	1/9,1/7,1/5
Kesinlikle Çok Düşük	KÇD	1/9,1/9,1/7

Adım 3: Uzman görüşlerinin, [28] tarafından önerilen bulanık yöntemle göre birleştirilmesi amacıyla Tablo 3’te verilen ölçek, s_0 en düşük ve s_g en yüksek aralığı göstermek üzere büyükten küçüğe doğru sıralanır. Böylece uzman görüşünün s_i ve s_j arasında değişmesi halinde, $s_0 \leq s_i < s_j \leq s_g$ ilişkisi geçerli olur. $\tilde{A} = (a, b, c)$ üçgen bulanık üyelik fonksiyonunun a ve c parametreleri (9) ve (11) formüllerinde tarif edildiği şekilde, b

parametresi ise OWA operatörü aracılığıyla şu şekilde hesaplanır:

$$b = \begin{cases} a_m^i & \text{Eğer } i + 1 = j \\ OWA_w(a_m^i, \dots, a_m^j) & \text{Aksi durumda} \end{cases} \quad (12)$$

OWA işleminde ihtiyaç duyulan ağırlık vektörü, [29] tarafından tarif edilmiştir: $[0,1]$ aralığında bir α parametresinin kullanıldığı, g 'nin Tablo 3'teki ölçekte yer alan terim sayısını, j 'nin en yüksek ve i 'nin en düşük değerlendirme değerini gösterdiği durumda $W = (w_1, w_2, \dots, w_n)$ şu şekilde hesaplanır:

$$w_1 = \alpha^{n-1}, w_2 = (1 - \alpha)\alpha^{n-2}, \dots, w_n = (1 - \alpha) \quad (13)$$

$$\alpha = (g-j+i) / (g-1) \quad (14)$$

Adım 4: $\hat{c}_{ij} = (c_{ij}^l, c_{ij}^m, c_{ij}^u)$ değerlerini içeren Ortak İkili Karşılaştırma Matrisi \hat{C} oluşturulur. Üçgen bulanık sayıların tersi, $\hat{c}_{ji} = (1 / c_{ij}^u, 1 / c_{ij}^m, 1 / c_{ij}^l)$ şeklinde hesaplanır.

Adım 5: $\hat{r}_i = (\hat{c}_{i1} \otimes \hat{c}_{i2} \dots \otimes \hat{c}_{in})^{1/n}$ formülü aracılığıyla; Ortak İkili Karşılaştırma Matrisindeki her satırın bulanık geometrik ortalaması tespit edilir.

Adım 6: $\hat{w}_i = \hat{r}_i \otimes (\hat{r}_1 \oplus \hat{r}_2 \dots \oplus \hat{r}_n)^{-1}$ formülü yardımıyla her kriterin bulanık ağırlık değeri elde edilir.

Adım 7: i kriterinin sıralamasının tespit edilmesi için, $D_i = w_{i,l} + 4w_{i,m} + w_{i,u} / 6$ eşitliğinden yola çıkarak üçgen bulanık sayılar, kesin sayılara dönüştürülür.

Adım 8: Bulanık olmayan ağırlık değerleri normalize edilir ve kriterler, kesin ağırlık değerlerine göre sıralanır.

4. KLASİK VE SEZGİSEL BULANIK İKİLİ KARŞILAŞTIRMA YÖNTEMİNİN UYGULANMASI (APPLICATION OF TRADITIONAL AND HESITANT FUZZY PAIRWISE COMPARISON)

Yazılım projelerinin maliyet tahmini için önemli kriterlerin tespit edilmesi amacıyla ilgili literatür detaylı incelenmiş ve uzman görüşünden yararlanmıştır [30, 31, 32, 33]. Böylece, bir yazılım geliştirirken sarf edilen çabanın tamamını kapsayacak şekilde en önemli kriterler geliştirilen ara yüz, uygulama, yığın iş, veri tabanı ve veri nesne sayısı olarak tespit edilmiştir. Uzmanlar ile yapılan yüz yüze görüşmeler ve doldurulan anketler aracılığıyla, maliyeti etkileyen bu kriterlerin bulanık olan ve olmayan ağırlık değerlerinin belirlenmesi için ihtiyaç duyulan bilgi toplanmıştır. Kriterlerin göreceli öneminin tespit edilmesi amacıyla, Türk yazılım sektöründe yönetici pozisyonunda çalışmakta olan 6 uzmanın görüşüne başvurulmuştur. Böylece 6 uzman; ara yüz (f_1), uygulama (f_2), yığın iş (f_3), veri tabanı (f_4) ve veri (f_5) nesne sayılarını temsil eden 5 farklı kriteri öncelikle Tablo 2'de verilen önem ölçüğü ile ikili olarak karşılaştırmıştır. Tablo 4, uzmanların Klasik İkili Karşılaştırma sonuçlarını göstermektedir.

Tablo 4: Maliyet Kriterlerinin Klasik İkili Karşılaştırması (Traditional Pairwise Comparison of Cost Criteria)

Uzman 1	f_1	f_2	f_3	f_4	f_5
f_1	1	3	5	9	9
f_2		1	1	5	7
f_3			1	1	5
f_4				1	5
f_5					1

Uzman 2	f_1	f_2	f_3	f_4	f_5
f_1	1	5	5	1	3
f_2		1	5	1	7
f_3			1	1	3
f_4				1	9
f_5					1

Uzman 3	f_1	f_2	f_3	f_4	f_5
f_1	1	3	3	1	9
f_2		1	7	7	5
f_3			1	5	3
f_4				1	9
f_5					1

Uzman 4	f_1	f_2	f_3	f_4	f_5
f_1	1	5	9	0,33	5
f_2		1	5	1	5
f_3			1	1	1
f_4				1	3
f_5					1

Uzman 5	f_1	f_2	f_3	f_4	f_5
f_1	1	2	7	5	9
f_2		1	3	9	3
f_3			1	1	1
f_4				1	7
f_5					1

Uzman 6	f_1	f_2	f_3	f_4	f_5
f_1	1	7	9	3	7
f_2		1	1	0,2	2
f_3			1	0,2	1
f_4				1	5
f_5					1

İkinci adımda; uzmanların Klasik İkili Karşılaştırma yoluyla bildirdiği görüşlerin geometrik ortalama ile birleştirilmesiyle elde edilen sonuçlar Tablo 5'te özetlenmiştir. Örnek olarak birinci kriterin ikinci kriterle kıyasla ortak ikili karşılaştırma değeri $(3 \times 5 \times 3 \times 5 \times 2 \times 7)^{(1/6)} = 3,83$ olarak hesaplanmıştır.

Tablo 5: Maliyet Kriterlerinin Klasik İkili Karşılaştırma Sonuçlarının Birleştirilmesi (Collaboration of Traditional Pairwise Comparison)

	f_1	f_2	f_3	f_4	f_5
f_1	1,00	3,83	5,91	1,88	6,52
f_2	0,26	1,00	2,84	1,99	4,41
f_3	0,17	0,35	1,00	1,00	1,89
f_4	0,53	0,50	1,00	1,00	5,91
f_5	0,15	0,22	0,53	0,17	1,00

Formül (2)'ye göre hesaplanan öz vektör ve kriterlerin ağırlık değerleri ise Tablo 6'da verilmiştir.

Tablo 6: Klasik İkili Karşılaştırma ile Elde Edilen Öz Vektör ve Ağırlık Değerleri (Eigenvector and Importance Weights Found by Traditional Pairwise Comparison)

	Öz Vektör Değerleri					Ağırlık Değerleri
	f_1	f_2	f_3	f_4	f_5	
f_1	0,47	0,65	0,52	0,31	0,33	0,46
f_2	0,12	0,17	0,25	0,33	0,22	0,22
f_3	0,08	0,06	0,09	0,17	0,10	0,10
f_4	0,25	0,08	0,09	0,17	0,30	0,18
f_5	0,07	0,04	0,05	0,03	0,05	0,05

Tablo 6'deki ağırlık değerleri incelendiğinde, yazılım geliştirme projelerinin maliyet tahmini üzerinde en büyük etkiye sahip kriterin %46'lık pay ile ara yüz ile ilgili nesne sayısını temsil eden kriter (f_1) olduğu görülmektedir. %22 ve %18 oranındaki ağırlığa sahip uygulama (f_2) ve yığın iş nesne sayısı (f_3) kriterlerinin de gerçekçi maliyet tahmini açısından önemli olduğu görülmektedir. Veri tabanı ve veri nesne sayısı ile ilgili kriterler (f_4 ve f_5) ise %10 ve %5 değerindeki ağırlık ile düşük öneme sahiptir.

Klasik İkili Karşılaştırma yönteminin yanı sıra, ayrıca belirsiz ortamlardaki durumun analiz edilmesi için uzmanlar; f_1, f_2, f_3, f_4 ve f_5 kriterlerini Tablo 3'te verilen bulanık dilsel ölçek yardımıyla karşılaştırmıştır. Tablo 7, 6 uzmana ait bulanık ikili karşılaştırma sonuçlarını göstermektedir.

Tablo 7: Maliyet Kriterlerinin Sezgisel Bulanık İkili Karşılaştırması (Hesitant Fuzzy Pairwise Comparison of Cost Criteria)

Uzman 1	f_1	f_2	f_3	f_4	f_5
f_1	TE	EDY	AY	ÇY	KÇY
f_2		TE	TE	ÇY	ÇY
f_3			TE	EDD	ZDY
f_4				TE	ÇY
f_5					TE

Uzman 2	f_1	f_2	f_3	f_4	f_5
f_1	TE	ZDY	ZDY	ZDD	EDY
f_2		TE	ÇY	EDD	AY
f_3			TE	EDY	EDY
f_4				TE	KÇY
f_5					TE

Uzman 3	f_1	f_2	f_3	f_4	f_5
f_1	TE	ZDY	EDY	TE	KÇY
f_2		TE	ÇY	ÇY	ÇY
f_3			TE	AY	ZDY
f_4				TE	ÇY
f_5					TE

Uzman 4	f_1	f_2	f_3	f_4	f_5
f_1	TE	AY	ÇY	AD	ÇY
f_2		TE	AY	EDY	ZDY
f_3			TE	ZDD	TE
f_4				TE	AY
f_5					TE

Uzman 5	f_1	f_2	f_3	f_4	f_5
f_1	TE	EDY	ÇY	ZDY	ÇY
f_2		TE	AY	KÇY	ZDY
f_3			TE	ZDD	TE
f_4				TE	AY
f_5					TE

Uzman 6	f_1	f_2	f_3	f_4	f_5
f_1	TE	AY	ÇY	ZDY	AY
f_2		TE	EDD	ZDD	EDY
f_3			TE	ÇD	ZDD
f_4				TE	ÇY
f_5					TE

Tablo 7'de yer alan değerlendirmeler incelendiğinde uzman görüşlerinin tutarlı olduğu, böylece söz konusu uzmanların görüşlerinden yararlanarak Ortak İkili Karşılaştırma Matrisinin oluşturulabileceği anlaşılmaktadır. Tüm uzmanların kriterleri değerlendirme aralığı, Tablo 8'de verilmiştir.

Tablo 8: Uzmanların Sezgisel Bulanık Değerlendirme Aralıkları (Intervals of Fuzzy Evaluations)

Uzman	f_1	f_2	f_3	f_4	f_5
f_1	TE	AY ve EDY arasında	ÇY ve EDY arasında	ÇY ve AD arasında	KÇY ve EDY arasında
f_2		TE	ÇY ve EDD arasında	KÇY ve ZDD arasında	ÇY ve EDY arasında
f_3			TE	ÇY ve AD arasında	ZDY ve ZDD arasında
f_4				TE	KÇY ve AY arasında
f_5					TE

(9), (10) ve (11) denklemleri aracılığıyla, üçgen bulanık kümeler belirlenmiştir. Tablo 8'de yer alan bulanık değerlendirmeler ve OWA operatörü kullanılarak bulunan üçgen bulanık kümeler Tablo 9'da sunulmaktadır. Örnek olarak; birinci kriterin ikinci ve üçüncü kritere kıyasla bulanık ağırlığının hesaplanma yöntemi şu şekildedir:

Birinci kriterin ikinci kritere kıyasla bulanık değeri AY ve EDY arasında, böylece $i = 6$ ve $j = 8$ değerindedir. (9) ve (11) denklemlerine göre, a ve c parametreleri $\{a_L^i, a_M^i, a_M^{i+1}, \dots, a_M^j, a_R^j\} = \{1, 3, 5, 7\}$ aralığının sırasıyla minimum ve maksimum değerlerini temsil etmektedir. Bu nedenle, $a = 1$ ve $c = 7$ 'dir. b parametresini hesaplamak için;

$$\alpha = (g-j+i) / (g-1) = (10-8+6) / (10-1) = 0,889$$

$$w_1 = \alpha^{2-1} = 0,889 \text{ ve } w_2 = (1 - \alpha)\alpha^{2-2} = (1-0,889) = 0,111 \text{ (} n = 2 \text{) olarak elde edilmiştir.}$$

$i + 1 \neq j$ olduğundan, (10) ve (12) denklemlerine göre $b = 0,889 \times 5 + 0,111 \times 3 = 4,778$ değerini alır. Böylece, birinci kriterin ikinci kritere göre bulanık değeri (1, 4,778, 7) şeklinde bulunmuştur.

Benzer şekilde birinci kriterin üçüncü kritere kıyasla bulanık değeri ÇY ve EDY arasında, böylece $i = 6$ ve $j = 9$ değerindedir. $\{a_L^i, a_M^i, a_M^{i+1}, \dots, a_M^j, a_R^j\} = \{1, 3, 5, 7, 9\}$ aralığının minimum ve maksimum değerleri olan a ve c parametreleri sırasıyla 1 ve 9 değerine sahiptir. b parametresini hesaplamak için;

$$\alpha = (g-j+i) / (g-1) = (10-9+6) / (10-1) = 0,778$$

$$w_1 = \alpha^{3-1} = 0,605, w_2 = (1 - \alpha)\alpha^{3-2} = (1-0,778) \times 0,778 = 0,173 \text{ ve } w_3 = (1 - \alpha) = 0,222 \text{ (} n = 3 \text{).}$$

$b = 0,605 \times 7 + 0,222 \times 5 + 0,173 \times 3 = 5,864$ değerine sahiptir. Böylece birinci kriterin üçüncü kritere göre bulanık değeri (1, 5,864, 9) olarak bulunmuştur.

Tablo 9: OWA Operatörü Aracılığıyla Elde Edilen Üçgen Bulanık Kümeler (Triangular Fuzzy Sets Found by OWA Operator)

	f_1	f_2	f_3	f_4	f_5
f_1	(1, 1, 1)	(1, 4,778, 7)	(1, 5,864, 9)	(0,14, 5,811, 9)	(1, 5,224, 9)
f_2	(0,143, 0,209, 1)	(1, 1, 1)	(0,33, 4,9, 9)	(0,2, 4,928, 9)	(1, 5,864, 9)
f_3	(0,111, 0,171, 1)	(0,111, 0,204, 3,03)	(1, 1, 1)	(0,11, 4,021, 7)	(0,2, 1,741, 5)
f_4	(0,111, 0,172, 7,143)	(0,111, 0,203, 5)	(0,143, 0,249, 9,091)	(1, 1, 1)	(3, 5,222, 9)
f_5	(0,111, 0,191, 1)	(0,111, 0,171, 1)	(0,2, 0,574, 5)	(0,111, 0,191, 0,333)	(1, 1, 1)

OWA operatörü yardımıyla Tablo 9'da yer alan tüm üçgen bulanık kümeler hesaplandıktan sonra, Bölüm 2.2.2'de tarif edilen yöntemin 5. adımına göre her satır için geometrik ortalama değerler bulunmuştur. Örnek olarak, birinci satırın geometrik ortalaması şu şekilde hesaplanır:

$$a_g = (1 \times 1 \times 1 \times 0,14 \times 1)^{1/5} = 0,675$$

$$b_g = (1 \times 4,778 \times 5,864 \times 5,811 \times 5,224)^{1/5} = 3,854$$

$$c_g = (1 \times 7 \times 9 \times 9 \times 9)^{1/5} = 5,515$$

Böylece, birinci satırın geometrik ortalaması (0,675, 3,854, 5,515) olarak elde edilmiştir. Daha sonra, üçgen bulanık kümede yer alan tüm değerler Bölüm 2.2.2'de sunulan yöntemin 6. adımına göre normalize edilmiştir. Geometrik ortalamaları normalize etmek için Tablo 3'teki ölçeğin en yüksek değeri (9) referans alınmış, böylece $(f_1 \oplus f_2 \oplus \dots \oplus f_n)$ çarpanı olarak 9 değeri kullanılmıştır. Bu durumda, birinci maliyet kriterine ait üçgen bulanık kümenin normalize değerleri şu şekildedir:

$$a_w = 0,675 / 9 = 0,075$$

$$b_w = 3,854 / 9 = 0,428$$

$$c_w = 5,515 / 9 = 0,613$$

Tüm kriterlere benzer işlem uygulanarak elde edilen bulanık ağırlık değerleri Tablo 10'da verilmiştir.

Tablo 10: Maliyet Kriterlerinin Bulanık Ağırlık Değerleri (Fuzzy Importance Weights of Cost Criteria)

Kriter	Bulanık Ağırlık Değerleri
f_1	(0,075, 0,428, 0,613)
f_2	(0,044, 0,219, 0,415)
f_3	(0,021, 0,084, 0,282)
f_4	(0,039, 0,060, 0,548)
f_5	(0,022, 0,036, 0,123)

Her maliyet kriterinin nihai önem düzeyinin elde edilmesi için, Bölüm 2.2.2'de tarif edilen yöntemin 7. adımına göre Tablo 10'da yer alan değerlerin bulanıklığı kaldırılmış ve değerler normalize edilmiştir. Örnek olarak, birinci kriter $D_1 = (0,075 + 4 \times 0,428 + 0,613) / 6 = 0,400$ şeklinde hesaplanmıştır.

Tablo 11: Maliyet Kriterlerinin Bulanık Olmayan ve Normalize Ağırlık Değerleri (Defuzzified and Normalized Importance Weights of Cost Criteria)

Kriter	Bulanık Olmayan Ağırlık Değerleri	Normalize (Net) Ağırlık Değerleri
f_1	0,400	0,437
f_2	0,223	0,243
f_3	0,107	0,117
f_4	0,138	0,151
f_5	0,048	0,052

Tablo 11'de görüldüğü üzere, ara yüz ile ilgili nesne sayısını temsil eden kriter (f_1), %43,7 oranındaki ağırlık değeri ile yazılım projelerinin maliyetini etkileyen en önemli etmendir. Uygulama nesne sayısı (f_2), ikinci en önemli kriter olup %24,3 önem düzeyine sahiptir. Yığın iş ve veri tabanı nesne sayısı ile ilgili kriterler (f_3 ve f_4), %11,7 ve %15,1 değer ile birbirine yakın önemdedir. Son olarak, veri obje sayısı (f_5) ile ilgili kriterin yazılım

projelerinin maliyet tahmini açısından çok düşük öneme (%5,2) sahip olduğu görülmüştür.

Önerilen Klasik ve Sezgisel Bulanık İkili Karşılaştırma yöntemleri, Türkiye'de hizmet vermekte olan büyük bir yazılım şirketinde uygulanmıştır. Bu amaçla öncelikle, işletmede 2015 yılında tamamlanan 1.180 adet yazılım geliştirme faaliyeti için f_1, f_2, f_3, f_4 ve f_5 kriterleri ile temsil edilen nesne sayılarına ait gerçek bilgiler toplanmıştır. Daha sonra, Tablo 6 ve 11'deki ağırlık değerleri kullanılarak Klasik ve Sezgisel Bulanık İkili Karşılaştırma yoluyla ilgili çalışmaların maliyeti tahmin edilmiştir. Sonuç olarak, önerilen yöntemlerin etkinliğinin incelenmesi için gerçekleşen ve tahmini maliyet arasındaki sapma incelenmiştir. Tablo 12, Klasik ve Sezgisel Bulanık İkili Karşılaştırma yöntemleri ile tahmin edilen maliyetlerin gerçek maliyetlere uzaklığını göstermektedir.

Table 12: Önerilen Yöntemler ile Elde Edilen Sonuçların Gerçekleşen Maliyet ile Kıyaslanması (Benchmark of Estimated Cost Found by Proposed Methods with Real Effort)

Gerçek Maliyet Aralığı (Adam-Gün)	Proje Sayısı	Ortalama Gerçek Maliyet (Adam-Gün)	Klasik İkili Karşılaştırma ile Ortalama Tahmini Maliyet (Adam-Gün)	Sezgisel Bulanık İkili Karşılaştırma ile Ortalama Tahmini Maliyet (Adam-Gün)	Klasik Yöntem Sapma (%)	Sezgisel Bulanık Yöntem Sapma (%)
0<x<10	164	6,74	5,98	6,06	11,28	10,09
10≤x<50	593	35,26	28,49	29,82	19,20	15,43
50≤x<100	358	64,08	74,85	70,95	-16,81	-10,72
x ≥ 100	65	192,32	226,09	218,04	-17,56	-13,37

Tablo 12'de yer alan sapma değerleri, $1 - (\text{Tahmini} / \text{Gerçekleşen Maliyet})$ formülü ile hesaplanmıştır. Bu nedenle pozitif sapma değerleri, bahsedilen aralıktaki ortalama tahmini maliyetin gerçekleşen maliyetten daha düşük olduğu anlamına gelmektedir. Tablo 12'de görüldüğü üzere, Klasik İkili Karşılaştırma ile tahmin edilen maliyet; gerçek maliyeti 0-10 ve 10-50 adam-gün arasında olan talepler için sırasıyla ortalama %11,28 ve %19,20 daha düşük iken gerçek maliyeti 50-100 adam-gün arasındaki ve 100 adam-günden daha fazla olan talepler için sırasıyla ortalama %16,81 ve %17,56 daha yüksektir. Öte yandan, Sezgisel Bulanık İkili Karşılaştırma ile tahmin edilen maliyet; gerçek maliyeti 0-10 ve 10-50 adam-gün arasında olan talepler için sırasıyla ortalama %10,09 ve %15,43 daha düşük iken gerçek maliyeti 50-100 adam-gün arasındaki ve 100 adam-günden daha fazla olan talepler için sırasıyla ortalama %10,72 ve %13,37 daha yüksektir. Tablo 12'deki verilere göre, Sezgisel Bulanık İkili Karşılaştırma yönteminin Klasik İkili Karşılaştırma yönteminden daha başarılı sonuç verdiği görülmektedir. Özellikle Sezgisel Bulanık İkili Karşılaştırma ile proje bazında elde edilen tahmini maliyet değerleri, ikili karşılaştırma sırasında görüş bildiren uzmanlar tarafından doğrulanmıştır.

5. SONUÇLAR VE GELECEK ÇALIŞMALAR (CONCLUSION AND FUTURE RESEARCHS)

Bu çalışmada, yazılım projelerinin maliyetinin tahmin edilmesine yönelik yöntemler önerilmiştir. Bölüm 1’de tarif edildiği üzere; yazılım maliyetinin kestirilmesi amacıyla literatürde yaygın olarak kullanılan tekniklerden analitik olan ve olmayan yöntemlerin bazı eksiklikleri bulunmaktadır. Örneğin COCOMO vb. analitik yöntemlerde, uzman görüşünden faydalanmamakta ve sonucun doğruluğu büyük ölçüde kullanılan maliyet kriterleri ile fonksiyona bağlıdır. Öte yandan, uzmanların yeterli deneyime sahip olmaması ya da objektif olamaması durumunda, uzman görüşü vb. analitik olmayan yöntemler ile tahmin edilen maliyetler de gerçekten uzaklaşmaktadır. Ayrıca, konu ile ilgili literatürdeki çalışmalar incelendiğinde analitik ve subjektif yaklaşımları birleştiren çalışmalara rastlanamamıştır.

Bu çalışmada, analitik ve nesnel yöntemler bir arada kullanılarak her iki yaklaşımın avantajlarından yararlanılmıştır. Ayrıca literatürde, yazılım projelerinin maliyetinin genellikle kesin ya da meta sezgisel yöntemler ile tahmin edildiği görülmüştür. Özellikle belirsiz ortamlarda maliyet tahminleme için uygun yöntem olarak çalışmada önerilen Sezgisel Bulanık İkili Karşılaştırma tekniğinin uygulama örneğine literatürde rastlanmamıştır. Bu nedenle, uzman görüşü ile belirlenen en uygun kriterlerin göreceli önemi, yine uzman görüşünden yararlanarak Klasik ve Sezgisel Bulanık İkili Karşılaştırma ile tespit edilmiş, Sezgisel Bulanık İkili Karşılaştırma sonuçlarının gerçek maliyete daha yakın olduğu görülmüştür. Elde edilen sonuçlar, görüş bildiren uzmanlar tarafından doğrulanmıştır.

Gelecekte, bu çalışmada tespit edilen ağırlık değerleri kullanılarak yazılım maliyetinin minimize edilmesinin amaçlandığı matematiksel modeller oluşturulabilir; atanabilecek en fazla sayıda kaynak, en uzun tamamlama süresi vb. kısıtlar ile problemin en iyi çözümü bulunabilir. Ayrıca proje sayısının çok fazla olduğu durumlarda, problemin en iyi çözümünün bulunabilmesi için sezgisel ve meta sezgisel yöntemler geliştirilebilir.

KAYNAKLAR (REFERENCES)

- [1] S.W. Lin, V.M. Bier, “A Study of Expert Overconfidence”, *Reliability Engineering & System Safety*, 93, 711-721, 2008.
- [2] H. Arkes, “Overconfidence in Judgmental Forecasting” in J. S. Armstrong, *Principles of Forecasting*, Kluwer Academic Publishers, 495-515, 2001.
- [3] A. Stellman, J. Greene, “Applied Software Project Management”, O’Reilly Media, 2005.

- [4] K. Wiegers, “Stop Promising Miracles”, *Software Development*, 8, 49–53, 2000.
- [5] E. Miranda. “Improving Subjective Estimates Using Paired Comparisons”, *Software, IEEE* 87-91, 2001.
- [6] G.N. Parkinson. “Parkinson’s Law and Other Studies in Administration”. Boston, USA: Houghton-Mifflin, 1957.
- [7] R. Nelson. “Management Handbook for the Estimation of Computer Programming Costs”, *Systems Development Corporation*, 1966.
- [8] C.E., Walston, C.P. Felix. “A Method of Programming Measurement and Estimation. *IBM Systems Journal*, 16, 54-73, 1977.
- [9] B.W. Boehm. “Software Engineering Economics”, Englewood Cliffs, NJ, Prentice-Hall, 1981.
- [10] B.W. Boehm. “Software Cost Estimation with COCOMO II. New Jersey, Prentice Hall, 2000.
- [11] S. Bilgaiyan, S. Mishra, M. Das, “A review of software cost estimation in agile software development using soft computing techniques”, *2016 2nd International Conference on Computational Intelligence and Networks (CINE)*, Bhubaneswar, 2016, 112-117.
- [12] R. Sarno, J. Sidabutar, Sarwosri, “Comparison of different neural network architectures for software cost estimation”, *2015 International Conference on Computer, Control, Informatics and its Applications (IC3INA)*, Bandung, 2015, 68-73.
- [13] S.M.S. Jafari, F. Ziaaddini, “Optimization of software cost estimation using harmony search algorithm”, *2016 1st Conference on Swarm Intelligence and Evolutionary Computation (CSIEC)*, Bam, 2016, 131-135.
- [14] F.S. Gharehchopogh, R. Rezaii, B. Arasteh, “A new approach by using tabu search and genetic algorithms in software cost estimation”, *2015 9th International Conference on Application of Information and Communication Technologies (AICT)*, Rostov on Don, 2015, 113-117.
- [15] F.S. Gharehchopogh, I. Maleki, A. Talebi, “Using hybrid model of artificial bee colony and genetic algorithms in software cost estimation”, *2015 9th International Conference on Application of Information and Communication Technologies (AICT)*, Rostov on Don, 2015, 102-106.
- [16] T.R. Benala, S.C. Dehuri, S.C. Satapathy, C.S. Raghavi, “Genetic algorithm for optimizing neural

- network based software cost estimation”, *International Conference on Swarm, Evolutionary, and Memetic Computing*. Springer Berlin Heidelberg, 2011.
- [17] N. Kushwaha, Suryakant, “Software cost estimation using the improved fuzzy logic framework”, *2014 Conference on IT in Business, Industry and Government (CSIBIG)*, Indore, 2014, 1-5.
- [18] A. Mittal, K. Parkash, H. Mittal, “Software cost estimation using fuzzy logic”, *ACM SIGSOFT Software Engineering Notes* 35, 1-7, 2010.
- [19] C.L. Hwang, K. Yoon. “Multiple Attributes Decision Making Methods and Applications”, Berlin Heidelberg, Springer, 1981.
- [20] T.L. Saaty. “The Analytic Hierarchy Process”, McGraw – Hill Inc. 1980.
- [21] J.J Buckley, “Fuzzy Hierarchical Analysis”, *Fuzzy Sets and Systems*, 17, 233-247, 1985.
- [22] B. Zhu, Z. Xu, “Analytic Hierarchy Process-Hesitant Group Decision Making”, *European Journal of Operational Research*, 239, 794– 801, 2014.
- [23] B. Oztaysi, S.C. Onar, E. Boltürk, C. Kahraman, “Hesitant Fuzzy Analytic Hierarchy Process”, *FUZZ-IEEE* , 1-7, 2015.
- [24] T.L., Saaty. “How to Make a Decision: The Analytic Hierarchy Process”, *European Journal of Operational Research*, 48, 9–26, 1990.
- [25] L. A. Zadeh, “Fuzzy Sets”, *Information and Control*, 8(3), 338–353, 1965.
- [26] F.T. Bozbura, A. Beskese, C. Kahraman, “Prioritization of Human Capital Measurement Indicators Using Fuzzy AHP”, *Expert Systems with Applications* 32, 1100-1112, 2007.
- [27] V. Torra, “Hesitant Fuzzy Sets”. *International Journal of Intelligent Systems*, 25, 529–539, 2010.
- [28] H. Liu, R.M. Rodriguez, “A Fuzzy Envelope for Hesitant Fuzzy Linguistic Term Set and Its Application to Multi-Criteria Decision Making”, *Information Sciences*, 258, 220–238, 2014.
- [29] D. Filev, R.R. Yager, “On the Issue of Obtaining OWA Operator Weights”, *Fuzzy Sets and Systems*, 94, 157-169, 1998.
- [30] I. Jacobson, “Object-Oriented Software Engineering: A Use Case Driven Approach”, **Addison-Wesley**, 1993.
- [31] K.. Ribu, “**Estimating Object-Oriented Software Projects with Use Cases**. MSc Thesis, Oslo, Norway, 2001.
- [32] H.M. Sneed, “Estimation of the Development Costs of Object-Oriented Software”, *Informatik Spektrum*, 19, 133-140, 1996.
- [33] T.W. Koh, M.H. Selamat, A.A.A. Ghani, R. Abdullah, “Review of Complexity Metrics for Object-Oriented Software Products”, *International Journal of Computer Science and Network Security*, 8, 314-320, 2008.

TinyOS Tabanlı Telsiz Duyarga Ağları için Bir Konumlandırma ve k -Bağlılık Denetleme Sistemi

Orhan DAĞDEVİREN¹, Vahid Khalilpour AKRAM²

^{1,2} Uluslararası Bilgisayar Enstitüsü, Ege Üniversitesi, İzmir, Türkiye
orhan.dagdeviren@ege.edu.tr, vahid59@gmail.com
 (Geliş/Received:09.06.2016; Kabul/Accepted:03.02.2017)
 DOI: 10.17671/gazibtd.309291

Özet— Telsiz duyarga ağlarında (TDAlarında) düğümlerin konumlarının bulunması ve aralarındaki bağlantıların denetimi önemli konulardandır. Konumlandırma probleminde, amaç konumları belli olmayan mobil veya statik düğümlerin koordinatlarını diğer kök düğümlerin yardımıyla bulmaktır. k -Bağlılık denetleme probleminde amaç en az kaç adet düğüm bozulduğunda ağın bağlılığının bozulduğunu bulmaktır. Bu çalışmada TDA için bir konumlandırma ve k -bağlılık denetleme sistemi tasarlanmıştır ve uygulanmıştır. Bu sistemde trilaterasyon algoritması kullanılarak düğümlerin güncel pozisyonları diğer kök düğümlerden gelen mesajların üzerinden bulunmaktadır. Uygulanan sistemde düğümlerin arasındaki mesafeler gelen sinyallerin RSSI (Received Signal Strength Indicator) değerinden tahmin edilir. Konumu belli olmayan bir düğüm, en az 3 kök düğümden mesaj aldıktan sonra kendi koordinatını hesaplayabilir. Tüm gönderilen mesajlar ve bulunan koordinatlar, ağın çıkış düğümü tarafından algılanıp, bir Java uygulamasına aktarılır ve arayüzler üzerinden kullanıcıya sunulur. Ağın son bağlılık durumunu göstermek için gelen mesajlar üzerinden oluşturulan güncel topoloji üzerinde bir k -bağlılık denetleme algoritması çalıştırılır. Önerilen ve MEMSIC-IRIS düğümlerin üzerinde denetlenen sistem tüm TinyOS işletim sistemini destekleyen düğümlerin üzerinde çalışabilir. Önerilen sistem çeşitli uygulamalarda bir altyapı olarak kullanılabilir.

Anahtar Kelimeler— Telsiz Duyarga Ağları, Konumlandırma, Trilaterasyon, Bağlılık Denetleme, k -Bağlılık

A Localization and k -Connectivity Detection System for TinyOS based Wireless Sensor Networks

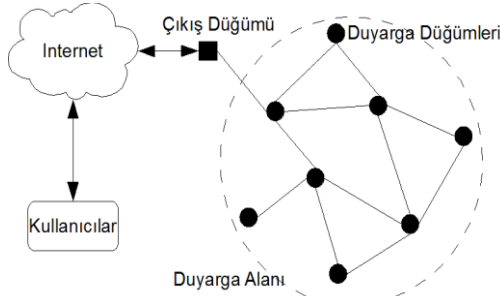
Abstract— Localization and connectivity detection are two important topics in Wireless Sensor Networks (WSNs). In localization problem the aim is finding the position of mobile or static nodes with unknown coordinates. The purpose of the k -connectivity problem is to find the number of nodes whose removal destroys the connectivity of the network. In this paper, we have designed and implemented a localization and k -connectivity detection system. The proposed system uses trilateration approach to find the up to date locations of nodes where the trilateration method uses the messages from anchor nodes. In our system, the RSSI (Received Signal Strength Indicator) values of transmitted messages are used to estimate the distance between nodes. Using the proposed system, nodes can estimate their locations after receiving messages from at least 3 anchor nodes. The exchanged message and detected positions are snooped by the sink node, transferred to a Java program and presented to the user on GUIs. To find the latest connectivity statuses of a network, a k -connectivity testing algorithm periodically is run on the current detected topology from incoming messages. The proposed system has been tested with MEMSIC-IRIS nodes and supports all WSNs which use TinyOS based nodes. This system can be used as an infrastructure for various applications in WSNs.

Keywords— Wireless Sensor Networks, Localization, Trilateration, Connectivity Detection, k -Connectivity

1. GİRİŞ (INTRODUCTION)

Günümüzdeki sistemler artık küçük ve ucuz bir entegre devre üzerinde, az güç kullanan haberleşme ünitesiyle tasarlanmaya başlanmıştır. Bu tasarım tekniği sayesinde telsiz duyurğa ağlarının (TDAlarının) kullanılması mümkün olmuştur. TDA, çevreden algılama yapan ve telsiz haberleşen düğümlerin sabit bir altyapı olmadan oluşturdukları bir ağıdır. Bir TDAda yüzlerce hatta binlerce duyurğa düğüm kendini yönetebilecek şekilde tasarsız bir yapıda kurulur [1][2]. Çevre gözetimi, askeri takip, sağlık hizmetleri ve akıllı tarım TDAYla yapılabilecek örnek uygulamalardır [3-5].

Şekil 1’de örnek bir TDA yapısı verilmiştir. Duyurğa düğümlerin çevreden topladığı veri diğer düğümler üzerinden çok zıplamalı (hop) olarak çıkış düğüme (sink node) ulaşır. Çıkış düğümü, diğer düğümlerden gelen verileri toplar ve kullanıcıya yönlendirmek üzere diğer ağ yapısına iletilen ağ geçiti (gateway) vazifesini görür. Çıkış düğümü şekilde görüldüğü üzere Internet üzerinden verilerini gönderebileceği gibi uydu teknolojisi ve hücresel ağ teknolojisi de kullanabilir. Şekilde çift taraflı oklarla da gösterildiği üzere bazı durumlarda kullanıcılar TDAYı değiştirmek için gerekli bilgileri gönderebilmektedir.



Şekil 1. Telsiz duyurğa ağlarının genel yapısı
(General architecture of wireless sensor networks)

TDAlar üzerinde düğümlerin konumlarının bulunması ve ağıın bağıllığının denetlenmesi önemli problemlerdir. TDAda konumlandırma işlemi, konumunu bilmeyen bir düğümün, konumunu bilen diğer düğümleri temel olarak konumunu bulması işlemidir [6-11,15-27]. Birçok durumda, TDA içindeki bir düğümün konumunu öğrenmesi gerekmektedir. Örneğin hedef takip (target tracking) ve olay tespit (event detection) uygulamalarında konumlandırma işleminin önemi çok büyüktür. Büyük ölçekli TDA uygulamalarında, her düğümün konumunun kullanıcı tarafından verilmesi çok zordur. Bunun yanında, GPS pahalı bir tercih olabilmektedir. Bu sebeplere dayanarak konumlandırma işlemi TDAda çok önemlidir [6].

Genel olarak TDAda düğümlerin arasında bir güvenilir iletişim altyapısı olmadığından dolayı, düğümlerin arasında bulunan bağlantılar tüm ağıın güvenilirliğini etkilemektedir. Bağılı bir ağıda bütün düğümler arasında

bir patika bulunur başka bir ifadeyle bağılı bir ağıda her düğüm diğer düğümlerin yardımıyla verilerini hedef düğüme ulaştırır. En kötü durumda bir düğümün bozulması ağı iki bağımsız parçaya bölüp, bazı düğümleri ulaşılmaz hale getirebilir. Bu sebepten ağıın bağıllığının periyodik olarak kontrol edilmesi önemlidir. Eğer bir ağıda herhangi iki düğümün arasında en az k bağımsız patika varsa o ağı k -bağılıdır. Bir k -bağılı ağıda herhangi $k-1$ düğümün bozulması, diğer düğümlerin arasındaki bağlantıları koparmaz. Bu nedenle güvenilir ağların k değeri yüksek olmalıdır. Ağıın k değerinin denetlenmesi önemli bir problemdir.

Bu çalışmada TDAlar için bir verimli konumlandırma ve k -bağıllık kontrol sisteminin tasarımı ve uygulanması anlatılmaktadır. Bu sistem IRIS düğümleri üzerinde TinyOS işletim sistemi için tasarlanıp, nesC ve Java dillerinde uygulanmıştır. Uygulanan sistemde TDAda bulunan düğümler kendi koordinatlarını diğer kök düğümlerin yardımıyla bulabilirler. Her düğüm koordinatını değiştirdiği zaman sistem otomatik olarak yeni konumunu hesaplar. Tasarlanan Java tabanlı arayüz sayesinde düğümlerin güncel konumları bulunur ve ağıın güncel bağıllık değeri kullanıcıya sunulur.

İlerleyen bölümler şu şekilde organize edilmiştir. Bölüm 2’de kullanılan mesafe ölçme yöntemi, konumlandırma yöntemi ve k -bağıllık denetleme yöntemi anlatılmıştır. Bölüm 3’de üzerinde çalışılan sistemin, tasarlanan TinyOS bileşenlerinin ve Java arayüzünün detayları verilmiştir. Bölüm 4’de sonuçlar değerlendirilmiştir.

2. KULLANILAN YÖNTEMLER (USED MATERIALS AND METHODS)

Bu bölümde tasarladığımız sistemde kullandığımız mesafe ölçme yöntemi, konumlandırma yöntemi ve bağıllık denetleme yönteminin açıklamaları verilmiştir.

2.1. Mesafe Ölçme Yöntemi (Distance Estimation Method)

TDAda düğümlerin arasındaki mesafeyi ölçmek için farklı yöntemler sunulmuştur ki bu yöntemlerin arasında Geliş Zamanına Göre Ölçme (Time of Arrival (TOA)) [7], Geliş Zamanı Farkına Göre Ölçme (Time Difference of Arrival (TDOA)) [8] ve Alınan Sinyal Gücüne Göre Ölçme (Received Signal Strength (RSS)) [9] bilinen yöntemlerdir. TOA yönteminde düğümlerin arasında zaman senkronizasyonu yapılması gereklidir. Zaman senkronizasyonu başlı başına önemli bir sorundur. TDOA yönteminde en az 2 alıcı/verici gereklidir ve Cricket gibi özel düğüm tasarımları bu donanıma sahiptir [10]. Bu sebeplerden dolayı bu çalışmada RSS yöntemi kullanılmıştır.

İki düğümün arasında mesafe ölçmenin basit yollarından birisi alınan sinyallerin gücüne bakmaktır. Radyo sinyalleri havada hareket ederken güçleri azalır ve bu yüzden göndericiye yakın bir düğümde alınan sinyal

gücü, uzaktaki düğümlere göre daha fazladır. Düğümlerde alınan sinyal gücü RSS olarak isimlendirirler ve dBm birimiyle ölçülür.

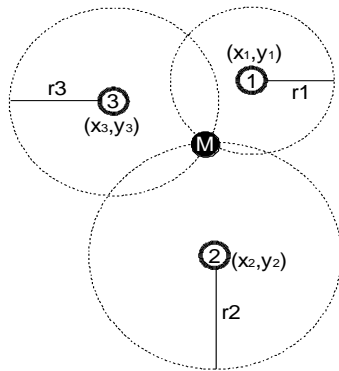
Yapılan denemelere göre genel olarak RSS değeri alıcı ve göndericinin uzaklığı ve aralarında bulunan engellere bağlı -40 dBm ve -140 dBm arasında olmaktadır. Eğer iki düğüm birbirine çok yakın olup, aralarında hiçbir engel olmazsa RSS değeri yaklaşık -40 dBm ve eğer iki düğümün arasındaki mesafe çok uzak veya aralarında bir nesne olursa RSS değeri yaklaşık -140 dBm olur. Genel olarak -140 dBm'den daha düşük alınan sinyaller parazit olarak gözüldüğüdür. Alınan RSS değerini kullanarak aşağıdaki formülden düğümlerin arasındaki mesafeyi tahmin edebiliriz [11]:

$$d = d_0 * 10^{\frac{RSS_0 - RSS}{10n}}$$

Yukarıdaki formülde d düğümlerin arasındaki mesafe, d_0 referans mesafe (örneğin 1 m), RSS_0 , iki düğümün arasındaki mesafe d_0 'iken alınan sinyal gücü, RSS gelen sinyalin gücü ve n sinyalin yolda kaybettiği güç oranıdır. n değeri farklı düğümlerde, antenin donanımı ve gücüne göre değişir.

2.2. Konumlandırma Yöntemi (Localization Method)

Trilaterasyon yönteminde konumlandırmanın yapılabilmesi için konumu belli olmayan bir düğümün, konumu bilen en az 3 adet kök düğümden, konum bilgilerini ve aralarındaki mesafeleri alması gerekmektedir.



Şekil 2. Trilaterasyon yöntemi (Trilateration method)

Şekil 2'de görüldüğü gibi üç çemberin kesişimi bir nokta olabilir ve böylece konumu belli olmayan düğümün koordinatları çember denklemlerinden bulunabilir. Her bir kök düğümün çemberi için aşağıdaki denklemi yazabiliriz:

$$(x_i - x)^2 - (y_i - y)^2 = r_i^2 \quad \text{for } i = 1,2,3 \quad (1)$$

Denklem 1'de (x, y) konumu belli olmayan düğümün koordinatı ve (x_i, y_i) , i nolu kök düğümün koordinatlarıdır. Eğer 3 nolu kök düğümün denklemini 1 ve 2 nolu kök düğümlerin denklemlerinden çıkarırsak, 2 ve 3 nolu denklemler elde edilir:

$$(x_1 - x)^2 - (x_3 - x)^2 + (y_1 - y)^2 - (y_3 - y)^2 = r_1^2 - r_3^2 \quad (2)$$

$$(x_2 - x)^2 - (x_3 - x)^2 + (y_2 - y)^2 - (y_3 - y)^2 = r_2^2 - r_3^2 \quad (3)$$

2 ve 3 nolu denklemleri (x, y) değişkenlerine göre düzenlersek aşağıdaki denklemler elde edilir:

$$2(x_3 - x_1)x + 2(y_3 - y_1)y = (r_1^2 - r_3^2) - (x_1^2 - x_3^2) - (y_1^2 - y_3^2) \quad (4)$$

$$2(x_3 - x_2)x + 2(y_3 - y_2)y = (r_2^2 - r_3^2) - (x_2^2 - x_3^2) - (y_2^2 - y_3^2) \quad (5)$$

4 ve 5 nolu denklemleri matris şeklinde yazabiliriz:

$$2 \begin{bmatrix} x_3 - x_1 & y_3 - y_1 \\ x_3 - x_2 & y_3 - y_2 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} (r_1^2 - r_3^2) - (x_1^2 - x_3^2) - (y_1^2 - y_3^2) \\ (r_2^2 - r_3^2) - (x_2^2 - x_3^2) - (y_2^2 - y_3^2) \end{bmatrix} \quad (6)$$

Böylece 6 nolu denklemden x ve y değişkenlerinin değeri diğer düğümlerin konumu ve mesafelerinden elde edilir. Örneğin bir TDAda kök düğümlerin koordinatlarının $(x_1, y_1) = (2,1)$, $(x_2, y_2) = (5,4)$, $(x_3, y_3) = (8,2)$ olduğunu varsayalım. Eğer koordinatı belli olmayan bir düğümün mesafesi bu kök düğümlere sırayla $r_1 = \sqrt{10}$, $r_2 = 2$ ve $r_3 = 3$ uzaktan olursa, bu düğümün koordinatı aşağıdaki şekilde bulunabilir:

$$2 \begin{bmatrix} 6 & 1 \\ 3 & -2 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 1 - (-60) - (-3) \\ -5 - (-39) - 12 \end{bmatrix} \quad (7)$$

$$\Rightarrow 6x + y = \frac{64}{2}, \quad 3x - 2y = \frac{22}{2} \Rightarrow (x, y) = (5,2)$$

Yukarıdaki denklemler bir doğrusal denklem sistemini oluşturmaktadırlar. Doğrusal denklem sistemlerini çözebilen algoritmalarından biri Gauss Jordan eleme yöntemidir. Genel olarak Gauss Jordan algoritması iki aşamadan oluşmaktadır. İleri yok etme (forward elimination) aşamasında $i > 1$ inci satırın tüm $j < i$ inci sütunlarındaki değerler bir önceki satırın değerlerinin katlarıyla toplanıp, bir üst üçgen matris elde edilir. İkinci aşamada elde edilen üst üçgen matrisinde geri yer değiştirme (backward substitution) yöntemiyle son satırdan başlayarak her değişkenin değeri bir önceki satırın değişkenlerinin aracıyla bulunur. Üst üçgen matrisinde son satırda sadece bir değer bulunduğu için, son değişkenin değeri direkt olarak bulunabilir. İleri yok etme yöntemi aşağıdaki algoritmada gösterilmiştir. İleri yok etme algoritmasında G matrisin referansı, n ise satır sayısıdır.

1: Algorithm Forward Elimination (G, n)

2: **Begin**

3: for $k=1$ to $n-1$ do

4: for $i=k$ to n do

5: $G_{ik} = G_{ik}/G_{kk}$

6: for $j=k+1$ to $n+1$ do

7: $G_{ij} = G_{ij} - G_{ik} * G_{kj}$

8: **End**

Yukarıdaki algoritma G matrisini bir üst üçgen matrisine çevirir ve ondan sonra değişkenlerin değeri aşağıdaki geri yer değiştirme algoritmasıyla bulunur. Geri yer değiştirme algoritmasında G matrisin referansı, n satır sayısı ve x yerine konulacak değişkendir.

1: Algorithm Backward Substitution (G, n, x)

2: **Begin**

3: for $i=n$ to 1 do

4: $x_i = G_{in+1}$

4: for $j=i+1$ to n do

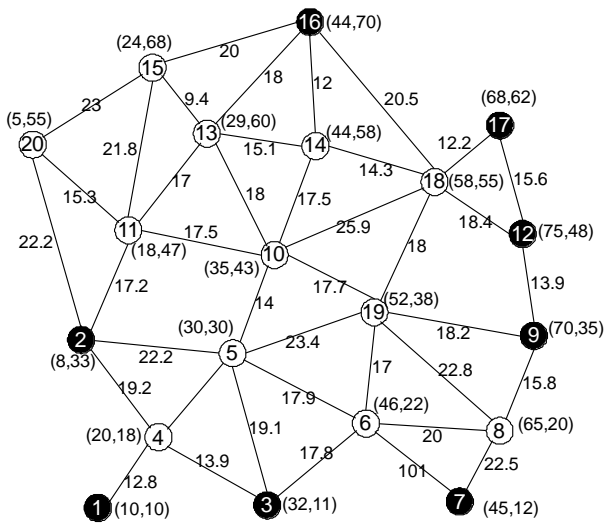
5: $x_i = x_i - G_{ij} * x_j$

6: $x_i = x_i/G_{ii}$

7: **End**

Gerçek yer değiştirme algoritmasında ilk olarak en son değişkenin değeri matrisin en son satırından bulunur ve böylece her değişkenin değeri bulunduktan sonra bir önceki satırdaki denklemde çözülebilir.

Şekil 3'de örnek bir TDA bulunmaktadır. Bu şekilde, kök düğümler siyah ve konumlarını bilmeyen düğümler beyaz renk ile gösterilmektedir. Kök düğümler, uygulamaya başlarken, kendi koordinatlarını tüm komşularına gönderirler. Beyaz düğümler, kendi koordinatlarını, siyah düğümlerden gelen mesajları kullanarak bulurlar ve buldukları koordinatları komşularına gönderirler. Böylece tüm beyaz düğümler kendi koordinatlarını bulabilirler.



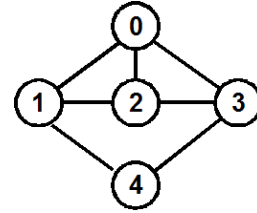
Şekil 3. Örnek bir duyurucu ağı
(A sample wireless sensor network)

Şekil 3'de kenarların üzerindeki sayılar düğümlerin arasındaki mesafelerdir. Ayrıca tüm düğümlerin

koordinatları onların yanında yazılmıştır. Bir siyah düğüm, kendi koordinatını hesaplamak için en az 3 komşusundan mesaj almalıdır. Bu nedenle Şekil 3'de, ilk aşamada, sadece 4 ve 18 numaralı düğümler kendi koordinatlarını hesaplayabilecekler. 4 numaralı düğüm kendi koordinatını hesaplandıktan sonra, bu koordinatı komşularına gönderir. Bu durumda, 5 numaralı düğüm 3 komşusundan mesaj alıp, kendi koordinatını hesaplar ve elde edilen bilgiyi yeni bir mesajda komşularına gönderir. Böylece sırayla 6, 8, 19, 10, 14, 13, 11, 15 ve 20 numaralı düğümler kendi koordinatlarını elde edebilirler.

2.3. k -Bağlılığı Denetleme Yöntemi (k -Connectivity Detection Method)

Bir ağın k -bağlılık değerini bulmak, o ağın güvenilirliğini belirleyen ana yollarından biridir ve bu yüzden k değerini denetleyen çeşitli algoritmalar sunulmuştur. Bir k -bağlı ağda aktif düğümlerin arasındaki bağlantıları tamamen kesilmesi için en az k düğümün bozulması gerekmektedir. Örneğin Şekil 4'de görünen çizgenin k değeri 2 olduğundan dolayı, bu çizgenin bağlantılığını bozmak için en az 2 düğümün silinmesi gerekmektedir. k -bağlılık denetleme algoritmalarının temel amacı mevcut bir ağın k değerini bulmaktır.



Şekil 4. Örnek bir 2-bağlı çizge
(A sample 2-connected graph)

Bu güne kadar yönsüz çizgelerde k değerini bulan asimptotik olarak zaman karmaşıklığı en iyi algoritma Henzinger tarafından sunulmuştur [12]. Bu sebeple bu çalışma bu algoritma üzerinden yapılmıştır. Henzinger, Hao [13] ve Even'in [14] minimum kesim ve k denetleme algoritmalarını birleştirerek, $O(kn^3)$ karmaşıklığına sahip bir denetim algoritması sunmuştur. *Henzinger-K-Detection* algoritması bu algoritmanın basamaklarını göstermektedir.

1: Algorithm Henzinger-K-Detection (G)

2: **Begin**

3: $i=0$;

4: $k=n-1$;

5: **repeat**

6: $i=i+1$;

7: $k=\min(k, \minCut(x_i, G), \minCut(x_i, G^R))$;

8: **until** $i>k$

9: **End**

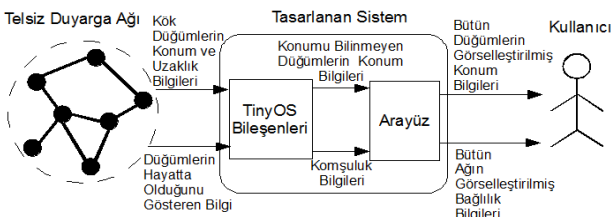
Henzinger-K-Detection algoritmasında \minCut fonksiyonu x_i düğümünü G 'den koparan en düşük düğümlerin kümesini vermektedir. Hao'nun algoritmasını kullanarak bu küme $O(n^3)$ karmaşıklığıyla bulunur. Herhangi bir k bağılı çizgede en fazla k düğümü

kontrol ettikten sonra, o çizgeyi bölen en düşük düğümlerin kümesi elde edilir. Eğer $C = \{c_1, c_2, \dots, c_k\}$ bir çizgeyi bölen düğümler kümesinin en küçüğü ise, herhangi $v \notin C$ düğümünün $minCut$ kümesi C' 'ye eşit olur. C kümesinin boyutu k olduğundan dolayı, çizgede en fazla $k + 1$ düğümün $minCut$ kümesini bulmak, k değerini belirler. Repeat-until döngüsünün en yüksek tekrarı k olduğu için, algoritmanın zaman karmaşıklığı $O(kn^3)$ olmaktadır.

3. TASARLANAN KONUMLANDIRMA VE k -BAĞLILIK DENETLEME SİSTEMİ (DEVELOPED LOCALIZATION AND k -CONNECTIVITY DETECTION SYSTEM)

TDAlar için tasarlanan konumlandırma ve k -bağlılık denetleme sistemi, iki temel bileşenden oluşmaktadır. İlk bileşen TinyOS işletim sistemini çalıştıran IRIS duyurga düğümleri için nesC dilinde hazırlanmıştır. Bu bileşenle düğümler kendi komşularını bulup, onlardan gelen mesajların sinyal gücü ve koordinatları üzerinden kendi koordinatlarını hesaplayabilirler. İkinci bileşen bir Java tabanlı arayüz yazılımı olarak, düğümlerin arasında gönderilen mesajları, ağda bulunan çıkış düğümünü aracılıyla algılayıp, düğümlerin güncel koordinatları ve ağın bağlantı durumunu denetler. Düğümlerin hareket etmesi durumunda değişen koordinatlar ve bağlantılar bu arayüz yazılımı tarafından kullanıcıya duyurulur. Bu bölümde ilk olarak üzerinde çalışılan donanım ve işletim sistemi anlatılacak, daha sonra tasarlanan TinyOS bileşenleri detaylı olarak açıklanacak, en son tasarlanan arayüz yazılımı incelenecektir.

Şekil 5'de tasarlanan sistemin genel yapısı gösterilmektedir. Bu şekilde görüldüğü gibi TDAda kök düğümlerin konum ve uzaklık bilgileri ve ayrıca diğer düğümlerin hayatta olduğu bilgileri çıkış düğümünü tarafından tasarlanan sisteme aktarılır. Bu sistem, gelen bilgileri girdi olarak alır, TinyOS bileşenlerini kullanır ve tüm düğümlerin koordinatları ve ağın bağlılık bilgileri olarak kullanıcıya arayüz üzerinden sunar.



Şekil 5. Tasarlanan sistemin genel yapısı
(General structure of designed system)

3.1 Üzerinde Çalışılan Donanım ve İşletim Sistemi (Underlying Hardware And Communication System)

Tasarlanan sistem IRIS düğümleri üzerinde çalıştırılmıştır. Şekil 6'da örnek bir IRIS düğümü gösterilmektedir. Bu düğümlerin özellikleri Tablo 1'de gösterilmektedir.



Şekil 6. Bir IRIS düğümü
(An IRIS node)

Tablo 1. IRIS cihazların özellikleri
(Main properties of IRIS nodes)

İşletim sistemi	TinyOS
Radyo Arayüzü	2.4 GHz, IEEE 802.15.4
İletim Hızı	250 kbps
İletim Gücü	3 dBm
Bellek	128 kB flash, 8 kB RAM

IRIS duyurga düğümleri üzerinde TinyOS işletim sistemi kullanılmıştır. TinyOS, TDAların düğümleri için düşük güç kullanarak TDAlar üzerinde uygulamaları daha kolay geliştirmeyi amaçlayan bir işletim sistemidir. TinyOS farklı bileşenlerden oluşmaktadır. Bu bileşenler uygulamalarda farklı amaçlar için kullanılmaktadırlar. Her uygulama kendi ihtiyaçlarına göre gereken bileşenleri programa ekleyerek daha az kaynak kullanarak amaçlarını gerçekleştirebilir. Genel olarak TinyOS işletim sistemi bir telsiz duyurga düğümünün bileşenlerine çeşitli servisler sunmaktadır.

Bir telsiz duyurga düğümü aynı zamanda farklı işlemler yapabileceğinden dolayı, bu düğümlerin işletim sistemi verimli bir şekilde eşzamanlı işlemleri desteklemelidir. Örneğin bir düğüm bir hesaplama yaparken, duyurga cihazından bir veri alma zorunda kalıp, aynı zamanda telsiz radyo üzerinden bir diğer düğümle iletişim kurabilir. Bu nedenle genel olarak TinyOS olay tabanlı bir işletim sistemidir. Uygulamalar her bileşenin farklı olayları için ayrı fonksiyonlar tanımlayabilirler ve böylece o olayın gerçekleşmesinde tanımlanan fonksiyon çalıştırılabilir. Örneğin cihazın açılması, radyo cihazından mesaj gelme, duyurgadan bilgi alma ve zamanlayıcı olayları, yaygın olarak kullanılan olaylardır.

TinyOS işletim sisteminin uygulamaları C dilinin gömülü sistemler için geliştirilmiş versiyonu, nesC diliyle geliştirilmektedir. nesC dilinde her uygulama çeşitli alt yapı bileşenlere bölünerek geliştirilir. Örneğin her düğümün radyo cihazı, ışıkları ve duyurgaları farklı bileşenlerle kontrol edilmektedir. Bu çalışmada tasarlanan sistem nesC diliyle yazılmıştır.

TinyOS işletim sistemine hazırlanan uygulama, TOSSIM benzetim ortamında az sayıda değişikliklerle çalıştırılabilir. TOSSIM, TinyOS işletim sistemine özel bir benzetim ortamıdır. nesC dilinde yazılan programlar TOSSIM benzetim ortamında, sanal topolojiler üzerinde çalıştırılabilir. Hazırlanan sistemde kullanıcı sanal bir topoloji oluşturarak, yazdığı kodunu, gerçek düğümler olmadan TOSSIM benzetim ortamı aracılığıyla çalıştırabilir.

3.2 Tasarlanan TinyOS Bileşenleri (Developed TinyOS Components)

Uygulanan sistemde düğümlerin arasında *Hello* ve *POS* mesajları gönderilir. *Hello* mesajları aracıyla düğümler kendi komşularını bulurlar. *POS* mesajları kök düğümler tarafından gönderilir ve diğer düğümler bu mesajları kullanarak kendi koordinatlarını hesaplarlar. *Hello* mesajı sayesinde düğümler komşuluk listelerini oluşturur. *POS* mesajının içinde göndericinin koordinatı bulunmaktadır. Önerilen sistemde gönderilen mesajların yapısı aşağıdaki gibidir.

```
typedef nx_struct MESSAGE {
    nx_uint8_t type;
    nx_uint8_t sender;
    nx_int16_t data;
    nx_int16_t X;
    nx_int16_t Y;
} MESSAGE_t;
```

Gönderilen tüm mesajlarda *type* mesajın tipi, *sender* mesajı gönderdiği düğüm, *data* mesajın tipine bağlı olarak bir bilgi, *X* göndericinin *x* koordinatı ve *Y* göndericinin *y* koordinatıdır.

TinyOS işletim sisteminde her bileşen farklı arayüzlerini kullanarak düğümün üzerinde bulunan cihazlardan yararlanabilir. Aşağıdaki kodda uyguladığımız TinyOS programında kullanılan arayüzleri gösterilmektedir.

```
module Algorithm @safe() {
    uses {
        interface Leds;
        interface Boot;
        interface Receive;
        interface AMSend;
        interface Timer<TMilli> as BroadcastTimer;
        interface SplitControl as AMControl;
        interface Packet;
        interface Random;
        interface Timer<TMilli> as Sender;
        interface PacketField<uint8_t> as PacketRSSI;
    }
}
```

Yukarıdaki arayüzleri düğümlerin radyo cihazı ve ışıklarının kullanmasını sağlayıp, aynı zamanda mesaj üretme, mesaj gönderme ve zamanlama işlemlerini de gerçekleştirmektelerdir. Bu arayüzler TinyOS işletim sisteminin uygun bileşenlerine bağlanarak cihazın donanımlar ve üzerinde çalışan yazılımlarının arasında bir köprü oluştururlar. Kullanılan TinyOS bileşenleri aşağıdaki kodlarda gösterilmektedir.

```
configuration AlgorithmApp {}
implementation {
    components MainC,Algorithm as App, LedsC,RandomC;
    components new AMSenderC(6);
    components new AMReceiverC(6);
    components new TimerMilliC() as T1;
    components new TimerMilliC() as T3;
    components RF230ActiveMessageC;
    components ActiveMessageC;
```

```
App.Boot -> MainC.Boot;
App.Receive -> AMReceiverC;
App.AMSend -> AMSenderC;
App.AMControl -> ActiveMessageC;
App.Leds -> LedsC;
App.Packet -> AMSenderC;
App.BroadcastTimer->T1;
App.Sender->T3;
App.Random->RandomC;
App.PacketRSSI -> RF230ActiveMessageC.PacketRSSI;
}
```

MainC.Boot bileşeni *Application.Boot* arayüzüne bağlanıp, düğüm çalıştırıldığı zaman *boot* fonksiyonunun çağırılmasına sebep olur. *Boot* fonksiyonu çağırıldığı zaman *Leds* bileşenin aracıyla düğümün üzerinde ışıkları açar ve *AMControl* bileşeninin *startDone* fonksiyonunu çağırarak cihazın radyo alıcı-göndericisini başlatır.

```
event void Boot.booted() {
    call Leds.set(1);
    call AMControl.start();
}
```

Yukarıdaki kodda *AMControl* bileşeninin *start* fonksiyonu çağırılmaktadır. Bu fonksiyon cihazın radyo alıcı-göndericisini çalıştırır. Cihazın radyo alıcı-göndericisinin başlatılması sona erdikten sonra aşağıda verilen *AMControl.startDone* fonksiyonu çağırılacaktır.

```
event void AMControl.startDone(error_t err) {
    if (err != SUCCESS) call AMControl.start();
    else {
        call Leds.set(0);
        call BroadcastTimer.startPeriodic(1000);
    }
}
```

Eğer radyo alıcı-gönderici başarıyla çalışmaya başlamazsa bu cihaz yeniden başlatılır. Aksi halde düğümün kırmızı ışığı açılarak *BroadcastTimer* isimli bir zamanlama bileşeni başlatılır. *startPeriodic* fonksiyonu bir zamanlayıcı planlayarak, her saniyede bir kere *fire* fonksiyonunu çağırır. Aşağıdaki kodda görüldüğü gibi *fire* fonksiyonu içinde pozisyonu belli olan düğümlerin koordinatları diğer düğümlere gönderilir.

```
event void BroadcastTimer.fired() {
    if (PosDetected)
        send(POS,mobCount,TOS_NODE_ID);
}
```

Yukarıda kullanılan *send* fonksiyonu mesajın içinde gönderilmesi gereken bilgileri aldıktan sonra, bir mesaj üretip, düğümün koordinatlarını mesaja ekleyerek, mesajı kendi komşularına gönderir.

```
void send(uint8_t typ,int16_t Data,uint8_t sender){
    int i=0;
    MESSAGE_t* out_msg;
    out_packet_size=sizeof(MESSAGE_t)+4*(Data-1);
    out_msg= (MESSAGE_t*)call Packet.getPayload(
```

```

    &out_packet,out_packet_size);
    out_msg->type=typ;
    out_msg->sender=sender;
    out_msg->data=Data;
    out_msg->X=myX;
    out_msg->Y=myY;
    for(i=0;i<Data;i++){
        out_msg->tag[i*2]=mobilesID[i];
        out_msg->tag[i*2+1]=mobilesDist[i];
    }
    i=call Random.rand16()%500;
    call Sender.startOneShot(i);
}

```

Düğümün arasında gönderilen mesajların çakışma ihtimalini azaltmak için *send* fonksiyonunda her düğüm 0 ve 500 milisaniye arasında rasgele bir sayı üretir ve o kadar bekledikten sonra mesajını gönderir. Bekleme işlemi bir diğer *Timer* bileşenin yardımıyla uygulanmaktadır. *send* fonksiyonunda *Sender* zamanlama bileşenin *startOneShot* fonksiyonu çağrılarak bu bileşenin *fire* fonksiyonu bir rasgele zamandan sonra çağrılır ve böylece mesaj gönderme işlemi bu fonksiyonun içinde gerçekleştirilir.

```

event void Sender.fired() {
    if(radioBusy==TRUE){
        return;
    }
    if (call AMSSend.send(AM_BROADCAST_ADDR,
        &out_packet, out_packet_size) == SUCCESS){
        radioBusy = TRUE;
        call Leds.led0Toggle();
    }
}
event message_t* Receive.receive(message_t* bufPtr,void* pkt,
uint8_t len){
    MESSAGE_t* payLoad;
    payLoad = (MESSAGE_t*) pkt;
    if(payLoad->type==POS){
        if(call PacketRSSI.isSet(bufPtr)){
            int d=call PacketRSSI.get(bufPtr);
            d=rssiToDistance(d);
            addToList(payLoad->sender,d, pkt->X, pkt->Y);
        }
        else addToNeighbors(payLoad->sender);
        return bufPtr;
    }
}

```

receive fonksiyonunda eğer gelen mesajın tipi *POS* ise mesajın RSSI değeri *PacketRSSI* bileşenin yardımıyla elde edilir ve dBm cinsinden olan değer mesafe cinsine çeviriler. Ondan sonra *addToList* fonksiyonu aracılığıyla göndericinin id'si, koordinatı ve mesafesi bir tabloya eklenir. Bu tablo düğümün pozisyonunu bulmak için trilaterasyon algoritması tarafından kullanılır. Eğer gelen mesajın tipi *POS* değilse mesajın göndericisinin id'si düğümün komşularının listesine eklenir.

Eğer yeni gelen mesajın göndericisinin bilgileri önceden tabloda bulunursa, *addToList* fonksiyonunda eski bilgiler yeni bilgilerle değişir. Aksi takdirde tabloya yeni bir satır eklenir.

```

void addToList(int16_t id,int16_t dist,int x, int y){
    int i=0;
    for(i=0;i<mobCount;i++){
        if(mobilesID[i]==id) break;
        nodeID[i]=id;
        nodeDist[i]=dist;
        nodeX[i]=x;
        nodeY[i]=y;
        if(i==mobCount)
            mobCount++;
    }
}

```

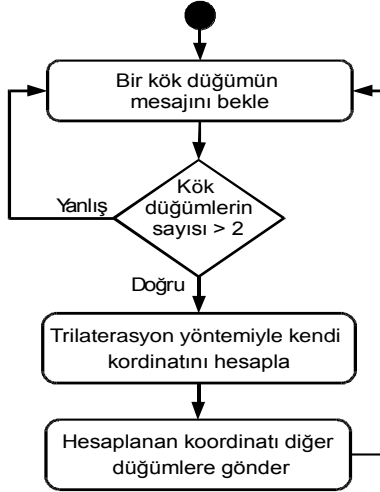
Tabloda en az 3 satır bulunduktan sonra düğümün koordinatı aşağıdaki fonksiyonda trilaterasyon yöntemi kullanarak hesaplanır ve diğer düğümlere gönderilir. Trilaterasyon algoritmasında elde edilen tablodaki değişkenlerin değeri Gauss Jordan yöntemiyle hesaplanmaktadır.

```

bool GSolve(double a[][3],uint16_t n,double *x)
{
    int i, j, k, max;
    double t;
    for (i = 0; i < n; ++i) {
        max = i;
        for (j = i + 1; j < n; ++j)
            if (a[j][i] > a[max][i])
                max = j;
        for (j = 0; j < n + 1; ++j) {
            t = a[max][j];
            a[max][j] = a[i][j];
            a[i][j] = t;
        }
        for (j = n; j >= i; --j)
            for (k = i + 1; k < n; ++k)
                a[k][j] -= a[k][i]/a[i][i] * a[i][j];
    }
    for (i = n - 1; i >= 0; --i) {
        a[i][n] = a[i][n] / a[i][i];
        a[i][i] = 1;
        for (j = i - 1; j >= 0; --j) {
            a[j][n] -= a[j][i] * a[i][n];
            a[j][i] = 0;
        }
    }
    for (j=0;j<n;j++)
        x[j] = a[j][n]/2;
    return(TRUE);
}

```

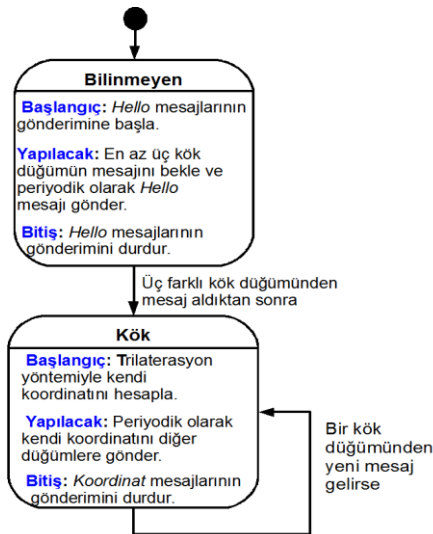
Pozisyonu belli olmayan düğümlerin üzerinde çalışan programın temel algoritması Şekil 7'de gösterilmektedir. Her düğüm ilk başta en az 3 kök düğümü tarafından *POS* mesajının gelmesini bekler. 3 *POS* mesajı geldikten sonra düğüm kendi koordinatını trilaterasyon yöntemiyle hesaplar ve koordinatını diğer düğümlere gönderir. Bu işlemden sonra döngünün başına dönerek yeni *POS* mesajlarının gelmesini bekler ve gelen her mesajın bilgilerine göre kendi koordinatını güncelleyip, yeni koordinatını diğer komşularına gönderir.



Şekil 7. Düğümlerin temel algoritması
(Base algorithm of nodes)

Bir Bilinmeyen durumunda bulunan düğüm en az 3 farklı kök düğümünden mesaj aldıktan sonra kendi koordinatını hesaplayıp ve kök durumuna geçer (Şekil 8). Kök durumundaki düğümler sürekli olarak diğer düğümlerden gelen mesajların yardımıyla kendi koordinatlarını güncelleyip, yeni koordinatlarını diğer düğümlere gönderirler.

Söz konusu olan programı gerçek düğümlerin üzerinde çalıştırılmadan önce benzetim ortamında da çalıştırabiliriz. TOSSIM benzetim ortamında bir benzetimi başlatmak için bir driver programı, C++ veya Python dilinde yazılabilir. Bu driver programı sanal bir topoloji oluşturarak, TinyOS programını çalıştırır ve topolojideki düğümlerin arasında, mesaj gönderimlerini gerçekleştirir. Şekil 3'deki topoloji için yapılan benzetimin çıktısı Şekil 9'da gösterilmektedir. Şekil 9'daki benzetim başlangıcında, 17 ve 7 numaralı düğümler, kök düğüm olarak, kendi koordinatlarını komşularına gönderirler.



Şekil 8. Düğümlerin durum diyagramı
(State diagram of nodes)

```

starting events..
DEBUG (17): Position broadcasted (68 62).
DEBUG (7): Position broadcasted (45 12).
DEBUG (18): Position 1 received from 17 (68,62).
DEBUG (8): Position 1 received from 7 (45,12).
DEBUG (6): Position 1 received from 7 (45,12).
DEBUG (2): Position broadcasted (8 33).
DEBUG (20): Position 1 received from 2 (8,33).
DEBUG (11): Position 1 received from 2 (8,33).
DEBUG (5): Position 1 received from 2 (8,33).
DEBUG (4): Position 1 received from 2 (8,33).
DEBUG (16): Position broadcasted (44 70).
DEBUG (15): Position 1 received from 16 (44,70).
DEBUG (13): Position 1 received from 16 (44,70).
DEBUG (14): Position 1 received from 16 (44,70).
DEBUG (18): Position 2 received from 16 (44,70).
DEBUG (1): Position broadcasted (10 10).
DEBUG (4): Position 2 received from 1 (10,10).
DEBUG (3): Position broadcasted (32 11).
DEBUG (6): Position 2 received from 3 (32,11).
DEBUG (5): Position 2 received from 3 (32,11).
DEBUG (4): Position 3 received from 3 (32,11).
DEBUG (4): X:20 Y:18
DEBUG (12): Position broadcasted (75 48).
DEBUG (9): Position 1 received from 12 (75,48).
DEBUG (18): Position 3 received from 12 (75,48).
DEBUG (18): X:58 Y:55
DEBUG (18): Position broadcasted (58 55).
DEBUG (14): Position 2 received from 18 (58,55).
DEBUG (10): Position 2 received from 5 (30,30).
DEBUG (19): Position 1 received from 18 (58,55).
DEBUG (4): Position broadcasted (20 18).
DEBUG (5): Position 3 received from 4 (20,18).
DEBUG (5): X:30 Y:30
DEBUG (5): Position broadcasted (30 30).
DEBUG (10): Position 2 received from 5 (30,30).
DEBUG (19): Position 2 received from 5 (30,30).
DEBUG (6): Position 3 received from 5 (30,30).
DEBUG (6): X:46 Y:22
  
```

Şekil 9. TOSSIM benzetim ortamının örnek çıktısı
(Sample screenshot of TOSSIM simulator)

Şekil 9'da 17 nolu düğümün koordinatını 18 nolu düğüm almıştır. 7 nolu düğümün koordinatını 8 ve 6 nolu düğümler almıştır. 2 nolu düğümün komşularına gönderdiği koordinat, 20, 11, 5 ve 4 nolu düğümler tarafından alınmıştır. 16 nolu düğümün komşularına gönderdiği koordinat, 15, 13, 14 ve 18 nolu düğümler tarafından alınmıştır. 1 nolu düğümün sadece bir komşusu olduğu için komşularına gönderdiği koordinat 4 nolu düğüm tarafından alınmıştır. Böylece 4 nolu düğüm ikinci mesajını almıştır. 3 nolu düğüm kendi koordinatını komşularına gönderir, 6 ve 5 nolu düğümler ikinci mesajlarını alacak, 4 nolu düğüm üçüncü mesajını alacaktır. 4 nolu düğüm üçüncü mesajını aldıktan sonra, kendi koordinatını bulur. Daha sonra 12 nolu düğüm kendi koordinatını komşularına gönderir. Geri kalan işlemler Şekil 8'de gösterildiği gibi benzer bir şekilde devam eder.

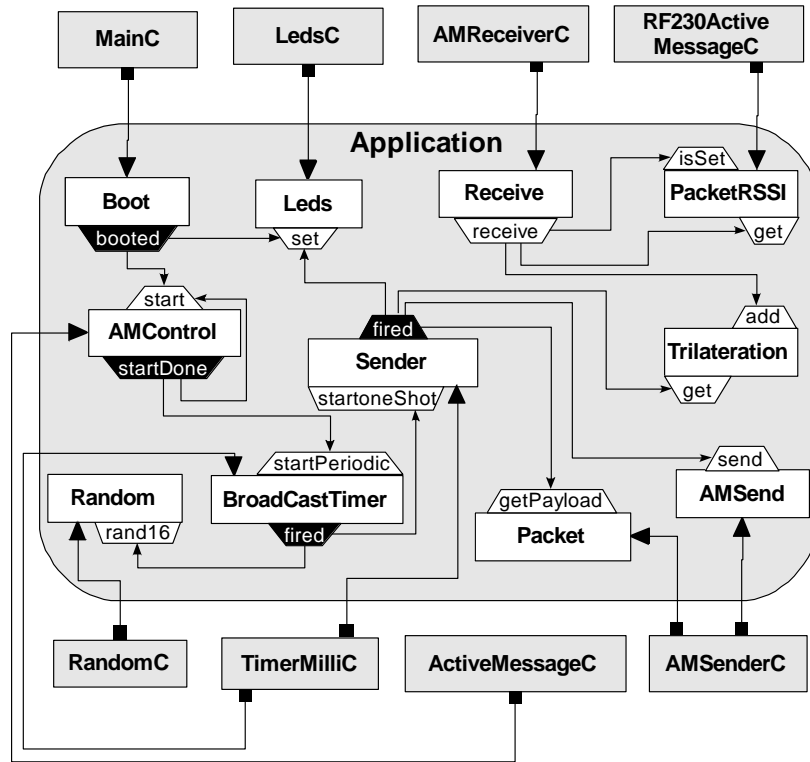
Uygulanan sistem, 15 adet MEMSIC-CROSSBOW IRIS düğümlerinin üzerinde test edilmiştir. Tüm kök düğümlerin koordinatları statik olarak programın içinde tanımlanmıştır. Sink düğümü bilgisayara bağlanarak, düğümlerin arasındaki iletilen mesajları dinleyip, bilgisayarda çalışan bir sonraki bölümde anlatılan Java uygulamasına gönderir. Şekil 10'da kullanılan düğümler ve sink düğümü gösterilmiştir.



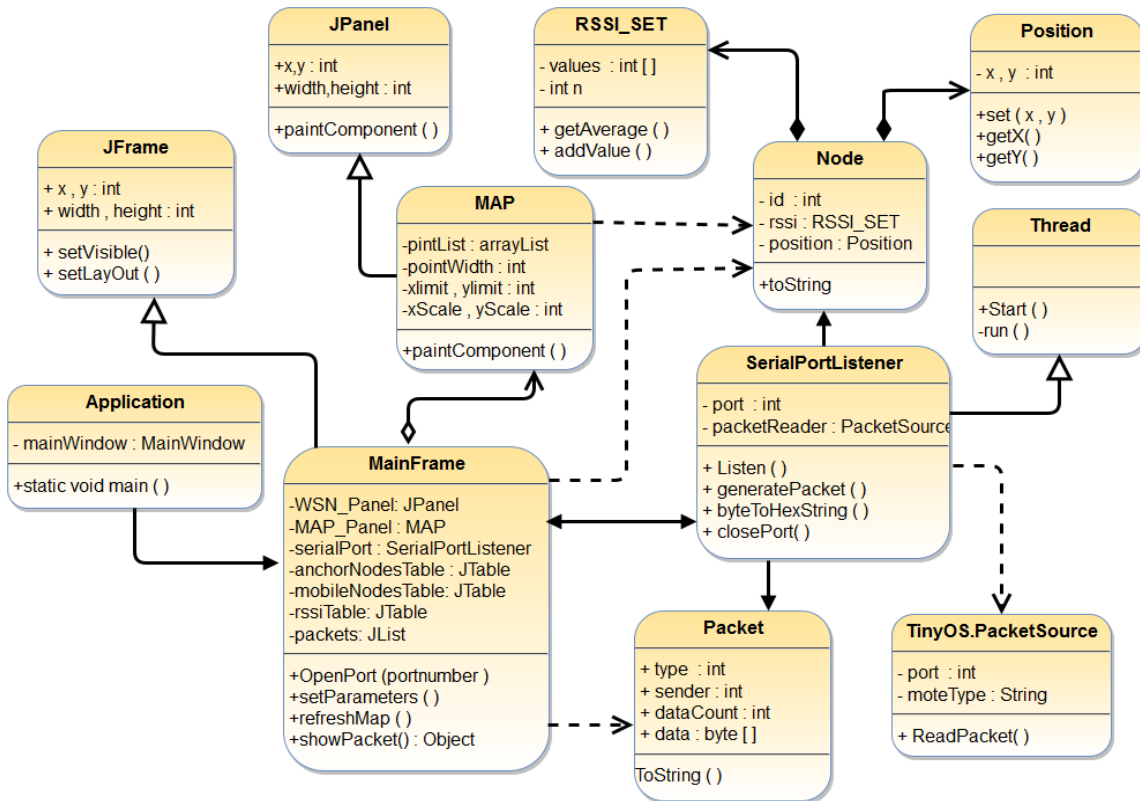
Şekil 10. IRIS düğümleri ve çıkış düğümü
(IRIS nodes and a sink node)

Uygulanan sistemin TinyOS bileşenlerinin ve onların arasında yapılan bağlantılar Şekil 11'de gösterilmektedir. Bu şekilde bileşenler gri renginde dikdörtgenlerle gösterilmektedir. Konumlandırma sistemine uyguladığımız *Application* isimli bileşen şeklin ortasında

çizilmiştir. *Application* bileşeninin etrafında görünen diğer gri dikdörtgenlerin hepsi TinyOS işletim sisteminde hazır bulunan bileşenlerdir ki *Application*'da kullanılan arayüzlere bağlanarak, uygulamanın düğüm üzerinde çalışmasını sağlamaktadırlar. Düğümlere gelen mesajlar *AMReceive* bileşeninin tarafından *Receive.receive* fonksiyonuna gönderilir. Bu fonksiyonda gelen mesajın tipine göre mesajın RSSI değeri *PacketRSSI* arayüzü aracılığıyla bulunur. IRIS düğümlerde gelen mesajların RSSI değeri *RF230ActiveMessage* bileşeni tarafından bulunmaktadır. Bu nedenle *PacketRSSI* arayüzü bu bileşene bağlanmıştır. *PacketRSSI*'da *isSet* ve *get* fonksiyonları bulunmaktadır. *isSet* gelen mesajda RSSI değerinin mevcut olmasını kontrol eder ve *get* fonksiyonu gelen mesajın RSSI değerini elde eder. Gelen *POS* mesajlarının RSSI değerleri elde edildikten sonra eğer en az 3 adet kök düğüm mesajı gelmişse *Trilateration.get* fonksiyonu çalıştırılır ve düğümün koordinatları hesaplanır.



Şekil 11. Uygulanan TinyOS programının bileşenleri
(Main components of implemented application)

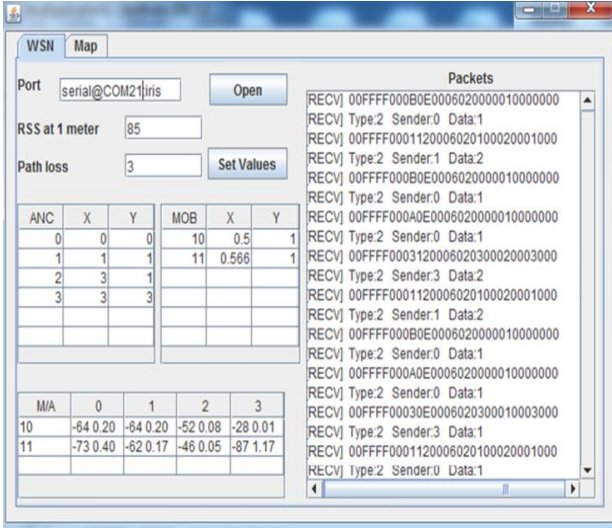


Şekil 12. Uygulanan Java arayüzünün sınıf diyagramı
(Main objects of implemented java interface)

3.3 Tasarlanan Arayüzü (Developed User Interfaces)

Düğümlerin gönderdiği mesajları, bulunduğu koordinatları ve ağın bağlılığını görmek için bir Java uygulaması geliştirilmiştir. Bu yazılım bilgisayara bağlanan çıkış düğümüyle USB port üzerinden iletişim kurarak ağda gönderilen tüm paketleri algılayıp, toplanan bilgileri uygun arayüzlerinde göstermektedir. Bu yazılımın sınıf diyagramını Şekil 12'de gösterilmektedir. Bu şekilde *MainFrame* sınıfı uygulamanın arayüzüdür. Bu sınıfın örnek çıktısı Şekil 13'de gösterilmektedir. Pencerenin sağ tarafında bulunan kutuda ağda gönderilen tüm paketler gösterilir. Bu kutuda paketlerin boyutları ve ayrıca içindeki veriler gösterilir. Sol tarafta bulunan tablolarda kök ve bilinmeyen düğümlerin id'leri, koordinatları ve birbirinden aldıkları RSSI değerleri gösterilmektedir. Kullanıcı girdi kutuları aracılıyla RSSI değerini mesafeye çevirme formülünde istediği değişiklikleri yapabilir. Pencerenin diğer sekmelerinde düğümlerin koordinatları ve aralarındaki bağlantılar harita ve çizge şekillerinde görünebilir.

MainFrame sınıfı ağda gönderilen paketleri algılamak için *PortListener* isimli bir sınıfı kullanmaktadır. *PortListener* sınıfı çıkış düğümünün port numarasını alır ve bu porttan gelen tüm verileri hem bayt ve hem paket olarak *MainFrame* sınıfına teslim eder. Bu sınıf *Thread* sınıfını genişleterek bir *Thread* olarak çalışır ve bu yüzden yazılım aynı zamanda hem gelen paketleri ve hem de kullanıcının komutlarını algılayabilir. Çıkış düğümü tarafından gönderilen paketler, *TinyOS.PacketSource* sınıfı aracılıyla porttan okunur ve gelen paketin bilgilerine göre bir *Node* veya bir *Packet* nesnesi üretilir. Bu nesnelere *MainFrame* sınıfına gönderildikten sonra gerekli tablolar güncellenir.



Şekil 13. Ağda gönderilen paketler
(Sent messages in the network)

```
public class PortListener extends Thread {
    private PacketSource reader;
    private JFrame1 window;
    private static final int AM_TYPE = 6;

    public PortListener(JFrame1 frm, String[] args) {
        window = frm;
    }
}
```

```
String source = null;
try {
    if (args.length == 2 && args[0].equals("-comm")) {
        source = args[1];
    }
    else if (args.length > 0) { System.exit(2); }
    if (source == null) {
        reader = BuildSource.makePacketSource();
    }
    else {
        reader = BuildSource.makePacketSource(source);
    }
    if (reader == null) {
        System.exit(2);
    }
    reader.open(PrintStreamMessenger.err);
} catch (Exception ex) {
    window.display(ex.toString());
}
}

public void Close() {
    this.stop();
    reader.close();
}

public void run() {
    try {
        Packet p;
        while (true) {
            byte[] packet = reader.readPacket();
            p = generatePacket(packet);
            window.display("[RECV] " + p.toString());
            window.processMessage(p);
        }
    }
    catch (Exception e) {
        window.display(e.toString());
    }
}
}
```

PortListener sınıfında, *PacketSource* sınıfından bir referans tanımlanmıştır. Gelen parametrelerde USB portunun numarası ve cihazın adı bulunmaktadır. Bu port *reader.open* fonksiyonunu çağırarak açılır. USB portu açıldıktan sonra *PortListener* nesnesi bir *Thread* olarak çalışmaya başlar. Bu sınıfın *run* fonksiyonunda sonsuz bir döngü içinde *reader.readPacket()* fonksiyonundan gelen mesaj, porttan bayt dizisi şeklinde alınır. Gelen mesaj *generatePacket* fonksiyonuyla *Packet* şekline dönüştürülür ve metin olarak *display* fonksiyonuyla ekrana yazılır. Ayrıca *processMessage* fonksiyonu, mesajı gönderen düğümün bilgilerini, tablolara ekledikten sonra, düğümü bir iki boyutlu alanda ve uygun koordinatta çizer.

Packet sınıfında düğümlerin bilgileri *MainFrame* sınıfına gönderilir. Her *Packet* nesnesinde gönderici düğümün id'si, düğümün koordinatı, mesajın tipi ve gereken ekstra bilgiler tutulmaktadır. Bu sınıfın yapısı aşağıdaki gibidir:

```
public class Packet {
    public byte type;
    public byte sender;
    public int data;
    public int X,Y;
    public Packet(){}
}
```

```

public Packet(byte type,byte sender,int data,int x,int y){
    this.type=type;
    this.sender=sender;
    this.data=data;
    this.X=x;
    this.Y=y;
}
}

```

MainFrame sınıfında mesajları işlemek için processMessage metodu yazılmıştır. Bu metod düğümlerin koordinatlarını alanda göstermek için refreshMap metodunu çağırır. Bu metodlar aşağıda verilmiştir.

```

public class MainFrame extends javax.swing.JFrame {

    PortListener serialPortGateWay;
    float RSSIAtOneMeter, pathloss;
    DrawGraph myMap;
    private HashMap RSSITable;
    public JFrame1() {
        initComponents();
        RSSIAtOneMeter=Float.parseFloat(jTextField3.getText());
        pathloss=Float.parseFloat(jTextField4.getText());
        RSSITable=new HashMap();
    }

    synchronized void processMessage(Packet pkt) {
        if (pkt.type == 2) {
            if (jTable1.getValueAt(pkt.sender,0) == null) {
                jTable1.setValueAt(pkt.sender, pkt.sender, 0);
                jTable1.setValueAt(pkt.X, pkt.sender, 1);
                jTable1.setValueAt(pkt.Y, pkt.sender, 2);
                jTable1.setValueAt(pkt.Z, pkt.sender, 3);
            }
            refreshMap ();
        }
    }

    public void refreshMap(){
        if (jTabbedPane1.getSelectedIndex() == 1) {
            jTabbedPane1.getSelectedComponent().repaint();
            this.pack();
        }
    }

    public void refreshGraph (JLabel label, short Mat[][] ) {
        int i, j;
        BeginDraw(false);
        for (i = 0; i < Mat.length; i++)
            for (j = i + 1; j < Mat.length; j++)
                if (Mat[i][j] == 1)
                    graph.add(i + " -> " + j + ";");
        EndDraw(label);
    }
}

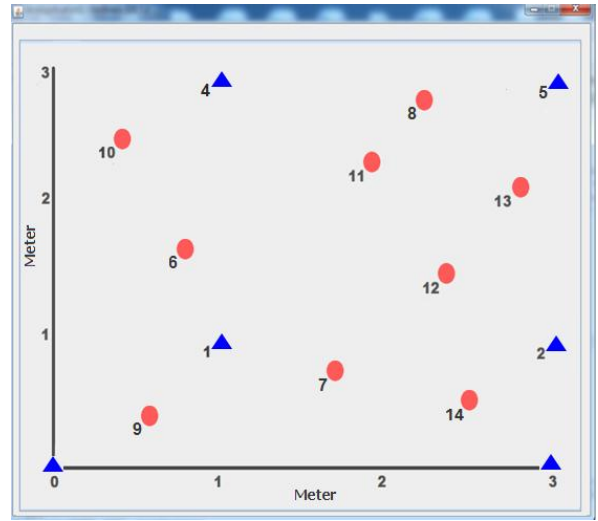
```

refreshGraph metodu tablolarda bulunan düğümleri çizge olarak kullanıcıya sunar. Bu metod düğümleri çizmek için GraphViz kütüphanesini kullanır. Bu metodun örnek çıktısı Şekil 14'de gösterilmektedir. Ağın k-bağlılık değerini ve ağdaki kritik düğümleri bulmak için Henzinger'in algoritması belli zaman aralıkları içinde çalıştırılır. Aşağıda bu işlemin kodu verilmiştir.

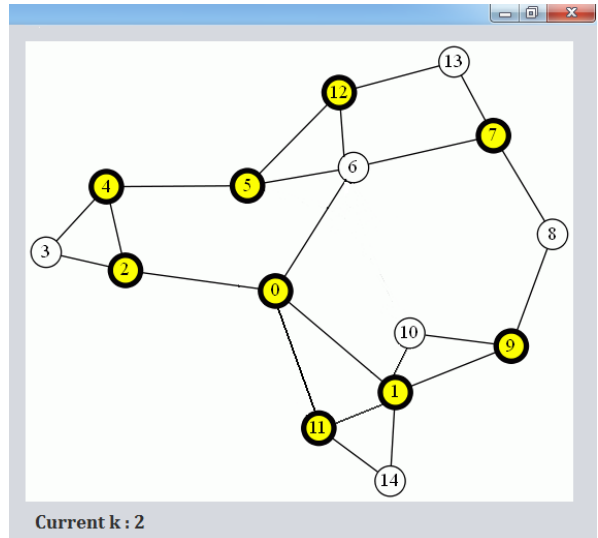
```

class Henzinger_K_Finder {
    private short[][] mat;
    public Henzinger_K_Finder (short[][] m) {
        mat = m;
    }
    public int Find_K() {
        int i=0, k=m.length()-1;
        do{
            i++;
            t=MinCut.find(m,i);
            if(t<k) k=t;
        }while(i>k);
        return k;
    }
}

```



Şekil 14. Düşümlerin koordinatları
(Coordinates of nodes)



Şekil 15. Ağın k-bağlı çizge olarak görüntüsü
(Monitoring network as k-connected graph)

Şekil 14'de haritanın X ve Y eksenlerinin limiti gelen paketlerin bilgilerinin üzerinden seçilir ve düğümler kendi

koordinatlarından çizilirler. Bu haritada kök düğümler mavi üçgenlerle ve koordinatlarını sonradan hesaplayan düğümler kırmızı dairelerle çizilmektedir. Koordinatı hesaplanmayan düğümler bu haritada çizilmemektedir ama bu düğümlerin id'leri diğer tablolarda görünebilir.

Şekil 15'de düğümlerin gönderdiği *Hello* mesajlarının üzerinden çizilen bağlantılar gösterilmektedir. Bu çizgede sarı ve daha kalın yuvarlaklarla çizilen düğümler bulunan k değeri için kritik düğümlerdir. Bu kritik düğümlerin bozulması veya hareket etmesi ağın k değerini 1 düşürebilir. Bu çizge çıktısında düğümler rasgele pozisyonlarda çizilmektedirler. Bu çizge sayesinde kullanıcıya ağdaki kritik düğümler ve ağın hata toleransı görsel olarak sunulur.

3.4 Konumlandırmanın Maliyeti ve Düğümlerin Enerji Tüketimi (Localization Overhead and Energy Consumption)

Tasarlanan sistemde kök düğümler, periyodik olarak koordinatlarını diğer düğümlere göndermektedirler. Konumunu bilmeyen her düğüm yeterli sayıda kök düğümden mesaj aldıktan sonra kendi koordinatlarını hesaplayabilir veya daha önce hesapladığı koordinatı düzeltir. Düğümler, carrier sense multiple access (CSMA) medium access layer (MAC) protokolünü kullanarak, paketlerini göndermeden önce 0 ile 500 ms arasında bir rasgele sayı çekip, bu sayı kadar bekledikten sonra paketlerini gönderirler. Bu protokol gönderilen paketlerin arasında bir çakışma ihtimaline yol açabilir. Ancak düğümler periyodik olarak mesajlarını rasgele zamanlarda gönderdiklerinden dolayı, ilerleyen periyotlarda yeni mesajlar alındığından düğümlerin konumlandırması gerçekleşir. Kök düğümlerde mesaj gönderimlerinin periyotları, düğümlerin enerji tüketimi ve konumlandırmanın hassasiyetini doğrudan etkilemektedir. Kısa gönderim periyotları, konum mesajlarının çakışma ve düğümlerin yer değiştirme etkisini azaltır ve hızlı hareket modellerinde bile daha güncel ve doğru konumlandırmaya yardımcı olur. Ancak, kısa süreli aralıkla mesaj gönderme, düğümlerin hızlı enerji tüketimine sebep olabilir.

Konumlandırma maliyetini periyodik olarak incelemek için düğümlerin enerji tüketimlerini kabaca hesaplayabiliriz. Duyarga düğümleri pille çalıştıklarından dolayı enerji tüketimleri önemli bir faktördür. Enerji tüketimini belirleyen en önemli etkenin radyo haberleşmesi olduğu bilinmektedir [8]. IRIS düğümlerinin veri sayfasına göre [28], 3.3 v güç kaynağıyla beslenen bu düğümler, 3 dBm radyo gücüyle çalışma durumunda, mesaj gönderirken 17 mA ve mesaj alırken 16 mA akı almaktadırlar. Bu düğümlerin iletişim oranı 250 kbps (31.25 kbps) olduğu için 8 baytlık konumlandırma mesajlarını yaklaşık 0.25 ms'de gönderilebilirler. Böylece, enerji tüketiminin genel ilişkisi $E = V * I * t$ 'yi dikkate alarak, bir konumlandırma mesajı, gönderici tarafta $3300 * 17 * 0.00025 = 14$ mj ve alıcı tarafta $3300 * 16 * 0.00025 = 13.2$ mj enerji tüketir. Dolayısıyla d adet komşusu olan bir düğüm her periyotta yaklaşık $14 + (d * 13.2)$ mj enerji tüketecektir. Bu

sebeple, mesaj gönderimlerinin periyotları, konumlandırmanın hassasiyeti ve aynı zamanda ağın toplam enerji tüketiminde önemli bir etkisi olduğundan dolayı bu sürelerin dengeli bir şekilde seçilmesi gerekmektedir. Uygulanan sistemde, kök düğümler her 20 s'de bir konum mesajı göndermektedirler. Düğümlerin hızlı hareket ettiği başka bir ağ için konumlandırma kalitesi çok önemli bir parametre ise periyot değeri düşük tutulmalıdır.

4. SONUÇLAR (CONCLUSIONS)

Bu çalışmada TDAlar için bir konumlandırma ve k -bağlılık denetleme sisteminin tasarımı ve uygulaması anlatılmıştır. Uygulanan sistemde düğümler ağda bulunan kök düğümlerin yardımıyla kendi koordinatlarını bulup, diğer düğümlerin konumlandırmasında yardımcı olur. Ayrıca kullanıcı hazırlanan Java yazılımıyla düğümlerin güncel pozisyonları ve ağın bağlantı durumunu öğrenebilir. Geliştirilen sistemde düğümlerin arasında mesafe ölçme işlemi, alınan mesajların RSSI değerinden ve lokalizasyon işlemi Trilaterasyon algoritmasıyla gerçekleştirilmektedir. Ağın k değeri gelen güncel mesajlar ve Henzinger algoritmasıyla sürekli kontrol edilir. Uygulanan tüm yazılım bileşenleri nesne tabanlı yöntemiyle geliştirildiği için, önerilen sistem bir altyapı olarak farklı amaçlara göre özelleştirilip, sisteme yeni işlevler eklenebilir.

TEŞEKKÜR (ACKNOWLEDGMENTS)

Yazarlar 113E470 numaralı proje kapsamında verilen araştırma desteğinden dolayı TÜBİTAK'a teşekkür eder.

KAYNAKLAR (REFERENCES)

- [1] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam and E. Cayirci, "A Survey on Sensor Networks", IEEE Communications Magazine, 40(8):102-114, 2002.
- [2] J. Yick, B. Mukherjee, and D. Ghosal, "Wireless sensor network survey", Comput. Netw. 52, 12 (August 2008), 2292-2330.
- [3] E. Biagioni and K. Bridges, "The Application of Remote Sensor Technology to Assist the Recovery of Rare and Endangered Species", Special issue on Distributed Sensor Networks for the International Journal of High Performance Computing Applications, 16(3): 315-324, 2002.
- [4] A. Mainwaring, J. Polastre, R. Szewczyk, D. Culler, and J. Anderson, "Wireless Sensor Networks for Habitat Monitoring", Proceedings of ACM International Workshop on Wireless Sensor Networks and Applications (WSNA'02), Atlanta, 88-97, 2002.
- [5] S. Dilek, and S. Özdemir, "Sağlık Hizmetleri Sektöründe Kablosuz Algılayıcı Ağları", international journal of informatics technologies, 7.2, 2014.
- [6] T. Alhmiedat and S. H. Yang, "A Survey: Localization and Tracking Mobile Targets through Wireless Sensors Network", PGNNet, 2007.
- [7] A. Jagoe, "Mobile location services: The definitive guide", Prentice Hall Professional, 1, 2003.

- [8] H. Karl, A. Willig, *Protocols and architectures for wireless sensor networks*, Wiley 2005, ISBN 978-0-470-09510-2.
- [9] J. C. Liberti and T. S. Rappaport, "Smart Antennas for Wireless Communications: IS-95 and Third Generation CDMA Applications", Upper Saddle River, Prentice-Hall, 1990.
- [10] N. B. Priyantha, A. K. Miu, H. Balakrishnan and S. Teller, "The cricket compass for context-aware mobile applications", *Proceedings of the 7th annual international conference on Mobile computing and networking*, 1-14, ACM, 2001.
- [11] J. Kang, D. Kim, Y. Kim, "RSS self-calibration protocol for WSN localization", *The 2nd International Symposium on Wireless Pervasive Computing ISWPC '07*, Puerto Rico, 2007.
- [12] M. Henzinger and G. N. Harold, "Computing vertex connectivity: New bounds from old techniques", *Journal of Algorithms*, 34.2: 222-250, 2000.
- [13] A. J. Hao and J. B. Orlin, "A faster algorithm for finding the minimum cut in a directed graph", *Journal of Algorithms*, 17: 424-446, 1994.
- [14] S. Even, "An algorithm for determining whether the connectivity of a graph is at least k ", *SIAM Journal on Computing*, 4(3): 393-396, 1975.
- [15] K. Langendoen, and N. Reijers, "Distributed Localization in Wireless Sensor Networks: A Quantative Comparison", *Elsevier Computer Networks*, 43: 499-518, 2003.
- [16] H. Lee, H. Dong, and H. Aghajan, "Robot-assisted Localization Techniques for Wireless Image Sensor Networks", *Proceedings of IEEE Conf. on Sensor, Mesh, and Ad Hoc Communications and Networks (SECON)*, 1: 383-392, 2006.
- [17] V. Ateş and M. A. Akcayol, "Kablosuz Ağlarda Tahmine Dayalı Hücreler Arası Geçiş Algoritmaları", *International journal of informatics technologies*, 3.3, 2010.
- [18] D. Niculescu, and B. Nath, "Ad-hoc positioning system", *Global Telecommunications Conference 2001, GLOBECOM'01*, IEEE, 5: 2926-2931, 2001.
- [19] P. N. Pathirana, N. Bulusu, A. V. Savkin, and S. Jha, "Node Localization Using Mobile Robots in Delay-Tolerant Sensor Networks", *IEEE Transactions on Mobile Computing*, 4(3):285-96, 2005.
- [20] C. Savarese, K. Langendoen, and J. Rabaey, "Robust positioning algorithms for distributed ad-hoc wireless sensor networks", *Proceedings of USENIX Technical Annual Conference*, 317-328, 2002.
- [21] A. Savvides, H. Park, and M. Srivastava, "The Bits and Flops of the N-hop Multilateration Primitive for Node Localization Problems", *Proceedings of First ACM International Workshop on Wireless Sensor Networks and Application (WSNA)*, 112-121, 2002.
- [22] Y. Shang, M. P. J. Fromherz, W. Ruml and Y. Zhang, "Localization from Mere Connectivity", *Proceedings of International Symposium on Mobile Ad Hoc Networking and Computing (MobiHoc)*, 201-212, 2003.
- [23] L. Chen, R. Yan, and Z. Ma, "TinyOS-based localization system design using accelerometer", *15th IEEE International Conference on Communication Technology (ICCT)*, IEEE, 511-518, 2013.
- [24] C. Chen, "Design of a child localization system on RFID and wireless sensor networks", *Journal of Sensors*, 2010.
- [25] N.P. Pathirana, B. Nirupama, V.S. Andrey and J. Sanjay, "Node localization using mobile robots in delay-tolerant sensor networks", *IEEE Transactions on Mobile Computing*, 4(3): 285-29, 2005.
- [26] J. Hill, R. Szewczyk, A. Woo, S. Hollar, D. Culler and K. Pister, "System Architecture Directions for Networked Sensors", *Proceedings of the ACM ninth international conference on Architectural support for programming*, 93-104, 2000.
- [27] L. Doherty, K. S. J. Pister and L. E. Ghaoui, "Convex Position Estimation in Wireless Sensor Networks", *Proceedings of IEEE INFOCOM*, 3: 1655-1663, 2001.
- [28] M. Inc, Iris datasheet, http://www.memscic.com/userfiles/files/datasheets/wsn/iris_datasheet.pdf.

Bulut Tabanlı Mobil Diyabet Kontrol Uygulaması: Mobil Diyabetim

Pelin YILDIRIM¹, Fatma BOZYİĞİT¹, Mehmet Hilal ÖZCANHAN², Semih UTKU²

¹Yazılım Mühendisliği Bölümü, Manisa Celal Bayar Üniversitesi, Manisa, Türkiye

²Bilgisayar Mühendisliği Bölümü, Dokuz Eylül Üniversitesi, İzmir, Türkiye

pelin.yildirim@cbu.edu.tr, fatma.bozyigit@cbu.edu.tr, hozcanhan@cs.deu.edu.tr, semih@cs.deu.edu.tr

(Geliş/Received:26.07.2016; Kabul/Accepted:11.12.2016)

DOI: 10.17671/gazibtd.309295

Özet— Yurdumuzda çok sayıda diyabet hastası bulunmaktadır. Bu hastalar diğer tüketiciler gibi tüm gün boyunca yoğun şekilde akıllı telefon ve mobil cihaz kullanmaktadır. Dolayısıyla, diyabet hastalarının kullandıkları mobil cihazlara özel bir yazılımla bu hastaların günlük yaşamlarını kolaylaştırmak mümkün görünmektedir. Bu çalışmada da diyabet hastalarına fiziksel veya tıbbi bir müdahalede bulunmadan, daha konforlu bir yaşam sürmelerine yardımcı olacak mobil bulut bilişim teknolojisine dayanan bir yazılım önerilmektedir. Hastaların diyabeti ilgilendiren günlük aktivitelerinin takibini kolaylaştıran özellikleriyle, geliştirilen çözümün hastaların iyileşme süreçlerine de yardımcı olacağına inanılmaktadır. Önerilen sistem birçok özellik yanında kullanıcılara özel, hem uyarıcı hem de hatırlatıcı mesajlar gönderebilmektedir. Benzer çözümlerle kıyaslandığında, önerilen yazılımın daha çok özelliği bir arada bulundurduğu ve iki yeni özellik içerdiği görülmektedir. Bir sonraki aşamada, çözümün yabancı dillere uyarlanması ve doktor gözetiminde gerçek denekler üzerinde teste tabi tutulması planlanmaktadır.

Anahtar Kelimeler— bulut bilişim, mobil bulut bilişim, diyabet, diyabet kontrol uygulaması

A Cloud Based Mobile Diabetes Control Application: My Mobile Diabetes

Abstract— There are a large number of diabetic patients, in our country. These patients use smart phones and mobile devices intensely throughout the day, as other consumers. Therefore, it seems possible to facilitate the daily lives of diabetes patients with special software for their mobile devices. In present study, medically non-intrusive, non-physical software based on mobile cloud computing that helps diabetic patients to have a more comfortable life is proposed. By facilitating the monitoring of patients' daily activities related to diabetes, it is believed that the developed solution can also help in mitigating some symptoms. Apart from multiple helpful features, the proposed system is capable of sending user specific warnings and alarms. When compared with similar solutions, the proposed system is observed to have more features in one solution and two new features. Future work is planned to translate the software into other languages and test the solution on real patients under doctor supervision.

Keywords— cloud computing, mobile cloud computing, diabetes, diabetes control application

1. GİRİŞ (INTRODUCTION)

Diyabet, insülin hormonu eksikliği sonucu ortaya çıkan ve henüz kesin bir tedavi yöntemi bulunmayan kronik bir hastalıktır. 2015 verilerine göre dünya genelinde yaklaşık olarak 415 milyon kişiye diyabet tanısı konulmuş ve bu hastaların 5 milyonunun hayatını kaybettiği belirtilmiştir[1]. Diyabet hastalığı kendi içerisinde Tip 1 (insüline-bağımlı diyabet) ve Tip 2 (insüline bağımlı olmayan diyabet) olmak üzere iki ana gruba ayrılmaktadır. Günümüzde diyabet hastalığının bakımı hastanelerde doktor muayenesi ve kan tahlili sonuçlarına göre uygun ilaçların verilmesi, sağlıklı beslenme ve düzenli egzersizlerin yapılması ile sağlanmaktadır. Ancak

nüfus yoğunluğu dolayısıyla, hasta sayısında sürekli artış olmakta ve hastanelerde diyabet hastalarının takibinin tümünde aynı başarının sağlanması mümkün olmamaktadır. Bazı diyabet hastalarının sağlık durumunun olumsuz seyretmesinin sebebi hastanelerdeki yoğunluktan dolayı tedavi ve tahlillerin gerektiği zamanda gerçekleştirilememesinden kaynaklanmaktadır. Buna hastaların randevularına ve diyabetle ilgili yapmaları gereken aktivitelere ilgi göstermemesi de eklenmektedir. Bu sebeplerden ötürü diyabet hastalarının tümünün sağlıklı beslenme, egzersiz yapma ve düzenli su tüketimleri düzene sokulamamaktadır. Bu durum, hastalık seyrinin kötüleşmesine ve ilave tıbbi müdahale,

gerektirebilmektedir. Sonuç olarak da olmaması gereken ek masraflar ve akut durumlar ortaya çıkmaktadır. Ülkemizde Sağlık Bakanlığı, diyabet hastalığı konusundaki maliyet yükünü kontrol altına almak için çalışmalarını hassasiyetle sürdürmektedir. 2013 yılında IDF tarafından yayınlanan 'Altıncı Diyabet Atlası' raporuna göre 2035 yılına kadar Türkiye dünyadaki diyabetli nüfus sayısının en yüksek olacağı ilk 10 ülke arasına girecektir [2]. Dolayısıyla, diyabet hastalarına günlük yaşamları sırasında yardımcı olabilecek kullanımı kolay, basit, her an yanlarında bulunabilecek araçlara ihtiyaç duyulmaktadır. Bu amaçla da, Sağlık Bakanlığı dijital teknolojilerin kullanımını diyabet hastalığı alanına da yaymayı ve personelinin diyabet hastalarına en iyi hizmeti sunabilmelerini hedeflemektedir.

Bulut bilişim ise günümüzün dijital teknolojisinin vardığı son noktadır. Bu teknoloji sanal bir sunucuda barındırılan uygulama ya da verilerin internet aracılığı ile her yerden erişilmesine olanak sağlayan büyük bir hizmetler bütünüdür. Bulut bilişim hizmeti müşterilerine otomatik ölçeklenebilirlik, kullandıkça öde seçeneği, her yerden erişilebilirlik, yüksek hacimli verilerin depolanması hizmetlerini sunmaktadır. En önemlisi ise, yüksek işlem gücüne sahip cihazlar ile başka yöntemlerle uzun süreye ihtiyaç duyulan veri analizinin, olabilecek en asgari sürede gerçekleştirilmesini sağlamaktadır.

Günümüzde, bulut bilişim altyapısının öneminin farkına varılarak sağlık alanındaki kullanımının arttığı gözlemlenmektedir. Bulut bilişim teknolojisi kullanılarak oluşturulan sistemlerde amaç hastalara ait verilerin etkin bir şekilde saklanması, sürekli erişilebilir olması, hızlı analizlerle sonuç alınarak, sağlık hizmeti performansının azami düzeyde tutulmasıdır. Böylece hastanelerde meydana gelen yoğunluktan kaynaklı tedavi, tahlil faaliyetlerinin gecikmesinden doğabilecek olumsuz sonuçların hafifletilmesi hedeflenmektedir. Özellikle, 6 milyon civarında diyabet hastasının mevcut olduğu ülkemizde, diyabet alanında verilen hizmeti rahatlatacak ve yersiz masrafları azaltacak sistemlere ihtiyaç duyulduğu açıkça görülmektedir.

Teknoloji geliştikçe mobil cihazlara verilen önem artmış ve cihazların kullanımı yaygınlaşmıştır. Akıllı telefon ve tablet bilgisayarlar gibi mobil cihazların özelliklerinde yaşanan gelişmeler mobil cihazların popülaritesini daha da arttırmıştır. Kullanım sayısında meydana gelen bu artış bulut bilişimin sağlamış olduğu ölçeklenebilirlik, yüksek işlem gücü, erişilebilirlik ve taşınabilirlik özelliklerinden dolayı bulut teknolojisinin kullanımına ihtiyaç doğurmuştur. Bunun sonucunda Mobil Bulut Bilişim (MBB) adı verilen bulut bilişim ile mobil cihazların birlikte çalışmasını sağlayan hizmetler bütünü kavramı ortaya çıkmıştır. MBB, mobil cihazların pil ömrünü uzatması ve genel işlem gücünü artırması, mobil cihazlara yüksek veri depolama alanı sunması, ölçeklenebilirlik gibi hizmetlerle birçok avantaj sunmaktadır. Hâlihazırda bazı telefon operatör şirketleri de MBB hizmetinden faydalanmaktadır. MBB'in yaygın olarak kullanıldığı sektörler ticaret, eğitim, sağlık ve oyun

endüstrisidir. Bu çalışmada, yukarıda tanıtılan teknolojileri kullanarak hastalarına evlerinde, kendi durumlarını takip ve kayıt altında tutabilecekleri MBB tabanlı bir diyabet kontrol çözümü önerilmektedir. Öneri Sağlık Bakanlığı'nın hedeflerine varmasına yardımcı olacak bir çalışmayı oluşturmaktadır.

Çalışmanın geri kalan bölümlerinde, ikinci bölümde ilk olarak benzer çalışmalar incelenmekte, kullanılan yöntemler ve veri setleri ile çalışma prensipleri hakkında bilgiler verilmektedir. Üçüncü bölümde önerilen sistem ve geliştirilen yazılım uygulaması hakkında detaylı bilgi verilmiştir. Dördüncü bölümde önerilen çözüm değerlendirilirken, beşinci bölümde ise sonuç ve gelecekte yapılması planlanan çalışmalardan bahsedilmektedir.

2. İLGİLİ ÇALIŞMALAR (RELATED WORKS)

Teknolojiyle birlikte sürekli olarak gelişen yazılım uygulamaları, sağlık alanında da oldukça yaygın bir şekilde kullanılmaktadır [3]. Sağlık alanında özellikle hastaların tedavilerini takip etmek, işlemlerini kolaylaştırmak ve pratik çözümler üretmek amacı ile birçok uygulama sunulmakta ve günden güne çeşitlilik ve kullanıcı yelpazesi büyümektedir. Özellikle, bulut bilişim sağlık alanında kullanılan uygulamalar içerisinde sıklıkla yer almaya başlamıştır. Örneğin, Hindistan'da yoğun nüfus sağlık alanındaki veri miktarında inanılmaz bir artışa sebep olmuş, bu da bulut bilişime olan ihtiyacı ortaya çıkarmıştır. PrincyMatlani ve Londhe sağlık hizmetini daha kaliteli ve etkin kullanmayı sağlamak adına, bulut bilişimin Hindistan'ın kırsal kesimlerinde tele-tıp alanında uygulanması üzerine bir çalışma gerçekleştirmiştir [4]. Bir diğer örnek, bir hastanenin pediatri ve doğum ile jinekoloji bölümlerinde kullanılmak üzere geliştirilen bilgi sistemidir [5]. Çalışmada hastanedeki her bir sağlık birimi için özel bulut bilişime dayanan altyapı oluşturulmuş ve hasta bilgilerinin birimler arası gönderilmesi için Health Level 7 (HL7) CDA standardını esas alan XML dosyaları kullanılmıştır. HL7 veri paylaşımında kullanılan standart kuralları barındırmaktadır. HL7 ile farklı ya da benzer ortamlardaki sağlık bilgi sistemlerinde bulunan verilerin belirli kurallar çerçevesinde paylaşılması hedeflenmektedir. Uygulamamızda farklı sistemlerin birbirleri ile uyumlu hale getirilmesi için HL7 kullanılmıştır.

2000'li yıllardan itibaren teknolojinin hızlı gelişiminden en büyük payı alan kablosuz iletişimde herkesin alabileceği bir konuma gelen mobil cihazlar büyük kitleler tarafından benimsenmiştir [6]. Mobil cihazlar için geliştirilen sağlıkla ilgili uygulamalar yaygın bir şekilde artmakta ve kullanılmaktadır. Mobil Sağlık Hizmeti (Mobile Healthcare) isimli bulut bilişim tabanlı mobil sağlık sistemi örneklerden biri olarak gösterilmektedir [7]. Çalışmada hastalara ait tıbbi dokümanların ve geçmiş bilgilerin, sistem içerisindeki paydaşlar ile eksiksiz ve sorunsuz bir şekilde paylaşıldığı iddia edilmektedir. İkinci olarak, Hoang ve Chen MoCASH isimli bulut bilişim tabanlı mobil destekleyici sağlık hizmeti altyapısı örnek

verilmektedir [8]. Araştırmacılar bu çalışmalarında bulut bilişimin avantajlarını sunarken, koordinat tabanlı altyapıların geliştirme zorluğu, ağ seçme ve mobil cihaz güç limitleri gibi dezavantajlarından da bahsetmişlerdir. Bu alanda verilebilecek kapsamlı örneklerden biri de Tayvan'da bulunan dört farklı hastaneyi kapsayan bir sağlık ağı oluşturma çalışmasıdır [9]. Çalışmada, hastalara ait tüm verileri tek bir bulut sisteminde depolayarak dört ayrı düğüm arasında dağıtım ve sorgulamayı kolaylaştıran bir sistem önerilmiştir. Manyetik rezonans ve röntgen gibi yüksek boyutlu verilere sahip hasta bilgilerinin sorunsuz şekilde depolanması ve hızlı şekilde sorgulanabilmesi mümkün hale geldiğinden, hastalar ve doktorlar arasındaki iletişimin arttırıldığı anlatılmaktadır. Çalışmanın bu yönleriyle sağlık hizmeti kalitesini arttırdığı da savunulmaktadır.

Son zamanlarda birçok araştırmacı diyabet hastalığının tedavisine yardımcı olmak amacı ile bulut altyapısı içeren mobil uygulamalar geliştirmeye başlamıştır. Bunlardan biri de diyabet tedavisinin kısa vizite süreleri, sağlık personeline erişimin zorluğu, hasta sayısının fazla olması ve bu nedenle hasta takiplerinin zor bir şekilde gerçekleştirilmesi gibi zorlukları ortadan kaldırmayı amaçlayan mobil diyabet kontrol uygulamasıdır [10]. Çalışmada, hastanın glisemi kontrolünü ve insülin dozajını ayarlamasını sağlayan bulut bilişim tabanlı bir sistem önerilmektedir. Uygulamayı kullanan hasta grubunun alyuvarlardaki hemoglobinin (Hb) kan şekere maruz kalması durumunda meydana gelen yapısal değişikliğin (glikozillenme) bir göstergesi olan HbA1c değerlerinin, kullanmayan gruba göre daha fazla düştüğü ve kullanıcılar tarafından olumlu tepkiler aldığı belirtilmektedir. Başka bir çalışmada Few Touch Application isimli bir mobil diyabet bakım uygulaması anlatılmaktadır [11]. Uygulamada hastanın kan şekeri verilerinin otomatik transferi, kısa hatırlatma mesajları gönderimi, diyabet günlüğü tutulması, fiziksel aktivite takibi gibi hizmetler sunulmaktadır. Uygulamalardan bir diğeri ise diyabet hastalarının glikozillenmiş hemoglobin seviyelerini düşürmeyi amaçlayan mobil takip uygulamasıdır [12]. Çalışmada, Maryland'da 163 diyabet hastası üzerinde yapılan araştırmalar ve uygulama ile hastaların glikozillenmiş hemoglobin seviyelerinde sağlanan düşüşler anlatılmaktadır. Farklı bir çalışmada da yetişkin diyabet hastalarının kan şekeri değerlerinin düşürülmesini hedefleyen bir mobil diyabet takip sistemi yer almaktadır [13]. Önerilen sistemle hastaların kan şekeri seviyelerinin geribildirimi, ilaç dozajlarının takibi yapılmaktadır. Çalışmalar sonucunda, uygulamayı kullananlarda kan şekeri değer düşüşü yaşandığı ve uygulamalarının hedefe ulaştığı bildirilmektedir.

Diyabet uygulamalarında genel olarak kullanıcıların şeker, tansiyon, öğün gibi bilgileri girdi olarak alınmakta ve değerler takip edilmektedir. Değerler belli eşiklerin üzerinde veya altında seyrettiği zaman hasta uyarılmaktadır.

3. ÖNERİLEN ÇÖZÜM (PROPOSED SOLUTION)

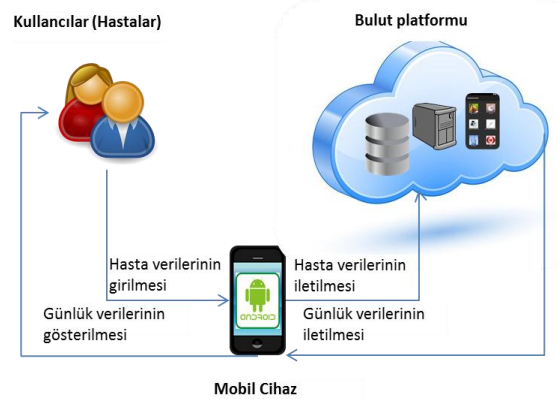
Önerilen çözüm diyabet hastalarının belirlenmiş saatlerde yapmaları gereken ölçüm ve faaliyetlerin (egzersiz yapmak ve su içmek gibi) zamanında yapılmasını sağlamaya yönelik hem sunucu hem de istemci yazılımlarından oluşan bir uygulamadır.

3.1. Motivasyon (Motivation)

Önerilen MBB destekli "Mobil Diyabetim" (MD) sistemi üç ana bölümden oluşmaktadır:

- Hastanın günlük tükettiği su miktarının hesaplanması ve günlük limitin altında su tüketilmesi durumunda hastaya hatırlatıcı mesajların gönderilmesi,
- Hastanın her öğün tüketmiş olduğu gıda verilerinin ve yapmış olduğu egzersizlerin sisteme girişinin sağlanması ve hasta tarafından bu verilerin sisteme girişinin unutulması durumunda belirli zaman aralıkları ile hatırlatıcı mesajların gönderilmesi,
- Ölçülen kan şekeri ve tansiyon değerlerinin sisteme kaydedilerek kan şekeri ve tansiyon günlüğü oluşturması, değerlerin sisteme girişinin unutulması durumunda belirli zaman aralıkları ile hatırlatıcı mesajların gönderilmesi.

MD sayesinde diyabet hastalarına ait veriler bulut sistemi üzerinde tutulacak ve etkin bir şekilde değerlendirme yapılması sağlanacaktır.



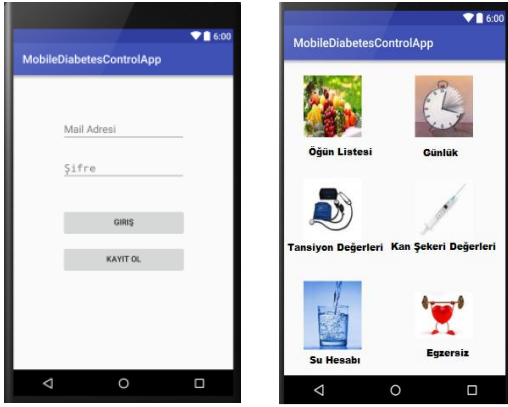
Şekil 1. Sistem mimarisi (System architecture)

3.2. Mobil Diyabet Yazılım Sistemi (Mobile Diabetes Software System)

Şekil 1'de, bulut tabanlı MD sistemin mimari yapısı gösterilmektedir. Sistem üç ana bölümden oluşmaktadır. Birinci bölümde, kullanıcıların (hastalar) gerekli tüm bilgileri (kişisel bilgiler, kan şekeri, tansiyon, su miktarı, öğün ve egzersiz gibi) mobil cihaza girmesi işlemi gerçekleştirilmektedir. İkinci bölümde, mobil cihaz üzerinden bilgilerin bulut platformuna iletilmesi, "Günlük" hizmeti ile de bulut platformunda tutulan bu bilgilerin tekrar kullanıcıya sunulması gerçekleştirilmektedir. Son bölümde ise, mobil cihaz aracılığıyla sayısız hasta

bilgilerinin yer ve zamandan bağımsız bulut bilişim platformu vasıtasıyla merkezi bir veri tabanında barındırılması sağlanmaktadır. Veri tabanında tutulan bilgilere, mobil uygulama tarafından her istenildiğinde hızlı bir şekilde erişilebilmektedir.

MD sadece sisteme daha önceden kayıtlı bulunan hastalar tarafından kullanılabilir bir uygulamadır. Bu sebeple, uygulama içerisinde kullanıcıyı ilk olarak Şekil 2 (a)'da gösterilen kullanıcı giriş ekranı karşılamaktadır. Önceden sisteme üye bir kullanıcı, mail adresi ve şifre bilgileriyle "Giriş" butonuna tıklayarak sisteme erişebilmektedir. Eğer kullanıcı sisteme daha önceden kayıt olmamış ise, "Kayıt Ol" butonu yardımıyla kayıt ekranına yönlendirilerek hastanın sisteme kayıt olması sağlanmaktadır. Hasta, isim soyadı, mail adresi, şifre ve telefon numarası bilgilerini girerek sisteme kayıt olmaktadır. Sisteme giriş yapan kullanıcı Şekil 2 (b)'de gösterilen Menü ekranına ulaşmaktadır. MD bu menü yardımı ile hastalara altı temel hizmet sunmaktadır. Sunulan hizmetler aşağıda detaylı şekilde anlatılmaktadır.



(a) (b)

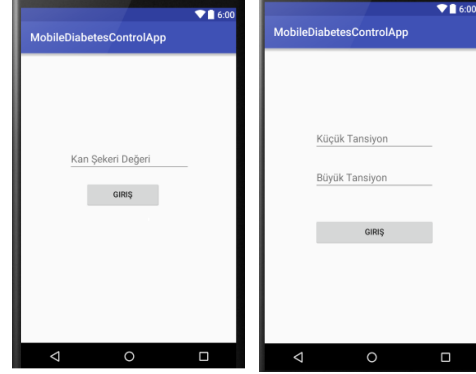
Şekil 2. (a) Kullanıcı giriş ekranı (User login screen)
(b) Menü ekranı (Menu screen)

3.2.1. Kan Şekeri Değerleri (Blood Glucose Values)

"Kan şekeri değerleri" butonuna tıklayarak kullanıcı Şekil 3 (a)'da gösterilen kan şekeri ekranına yönlendirilmektedir. Bu ekranda hasta ölçmüş olduğu kan şekeri değerlerini sisteme girmektedir. Hasta tarafından değerlerin sisteme girişinin unutulması durumunda belirli zaman aralıkları ile uygulama tarafından hastaya hatırlatıcı mesajlar gönderilmektedir. Hasta bilgilerini girdiği zaman mesaj gönderimi sona ermekte, ancak boş geçen bilgi girişleri veri tabanına "bilgi girilmedi" şeklinde yansıtılmaktadır.

3.2.2. Tansiyon Değerleri (Blood Pressure Values)

"Tansiyon değerleri" butonuna tıklayarak kullanıcı ise Şekil 3 (b)'de gösterilen tansiyon değerleri ekranına yönlendirilmektedir. Hasta ölçmüş olduğu büyük ve küçük tansiyon değerlerini bu ekran yardımı ile sisteme kaydetmektedir. Kan şekeri değerleri ekranında olduğu gibi, bu ekranda da hasta tarafından değerlerin sisteme girişinin unutulması durumunda belirli zaman aralıkları ile hatırlatıcı mesajlar gönderilmektedir.

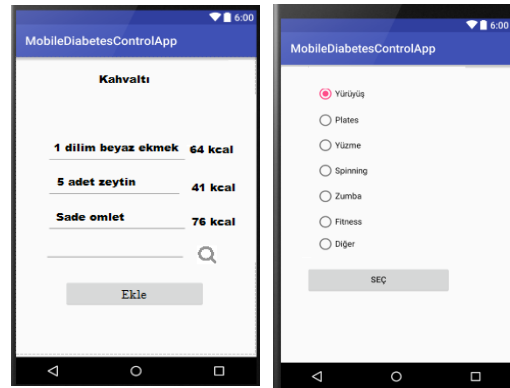


(a) (b)

Şekil 3. (a) Kan şekeri ekranı (Blood glucose screen)
(b) Tansiyon değerleri ekranı (Blood pressure screen)

3.2.3. Öğün Listesi (Meal List)

Kullanıcı "Öğün Listesi" butonuna tıklayarak Şekil 4 (a)'daki öğün listesi ekranına ulaşmaktadır. Kullanıcı bu ekran üzerinde seçmiş olduğu öğün türüne göre tükettiği besinleri girerek sisteme kaydetmektedir. Kullanıcının öğün bilgilerini girmeyi unuttuğu durumlarda da yine hatırlatıcı mesajlar ile kullanıcı uyarılmaktadır.



(a) (b)

Şekil 4. (a) Öğün listesi ekranı (Meal list screen)
(b) Egzersiz ekranı (Exercise screen)

3.2.4. Egzersiz Bilgileri (Exercise Information)

Hasta “Egzersiz” butonuna tıklayarak Şekil 4 (b)’deki egzersiz ekranına yönlendirilmektedir. Kullanıcı liste içerisindeki egzersiz türünü seçip, yapmış olduğu set sayısını girerek sisteme kaydetmektedir. Kullanıcı yetersiz egzersiz yapması veya egzersiz bilgilerini girmeyi unuttuğu durumlarda da yine hatırlatıcı mesajlar ile uyarılmaktadır.

3.2.5. Su Tüketimi Takibi (Water Consumption Tracking)

Hasta “Su hesabı” butonuna tıklayarak Şekil 5 (a)’da gösterilen su tüketimi ekranı üzerinden günlük tüketilen su miktarını bardak cinsinden girerek sisteme kaydedebilmektedir. MD girilen su bilgilerini belirli aralıklarla kontrol edip hastanın yeterli miktarda su tüketmemesi veya bilgileri sisteme girmeyi unuttuğu durumlarda hastaya hatırlatıcı mesajlar göndermektedir.

Görüldüğü gibi hasta birçok hatırlatıcı mesaj almakta, uyarılara uyması durumunda da hastalığının seyrini olumlu yönde etkileyecek gerekli tedbirleri yerine getirmektedir.

3.2.6 Günlük (Diary)

“Günlük” butonuna tıklayarak kullanıcı, Şekil 5 (b)’de gösterilen günlük ekranına yönlendirilir. Bu ekran üzerinde daha önceden sisteme girmiş olduğu tüm değerleri (kan şekeri, tansiyon, su miktarı, kalori ve egzersiz bilgileri gibi) gün bazında listeleyebilmektedir.



(a)

(b)

Şekil 5. (a) Su tüketimi ekranı (Water consumption screen)
(b) Günlük ekranı (Diary screen)

Günlüğünü kontrol eden ve MD uygulamasından ayrılmak isteyen kullanıcı “Çıkış” butonuna tıklayarak

sistemden güvenli bir şekilde çıkışını gerçekleştirebilmektedir.

4. DEĞERLENDİRME (EVALUATION)

Diyabet hastalarının hayatlarını kolaylaştırmak ve iyileşme süreçlerine yardımcı olmak amacıyla önerilen çözümlerin değerlendirilmelerinde öncelikle sınıflandırmaları yapılmaktadır. Çözüm önerilerinin sınıflandırmasında diyabet takip uygulaması olup olmadıkları, mobil cihazlar için tasarlanmış olup olmadıkları ve bulut bilişim altyapısı kullanmayı önerip önermedikleri değerlendirilmektedir. Tablo 1’de, benzer çalışmalar ve MD önerisinin bu üç konudaki durumları kıyaslanmaktadır. Tablo’da bulut bilişimi sağlık hizmetleri alanında öneren çalışmalara yer verilmiştir. Ancak, incelenen çalışmalardan sadece [10], [11], [12], [13] ve MD çalışmalarının MBB tabanlı mobil diyabet takip uygulaması oldukları görülmektedir.

Sundukları hizmetler açısından da önerilen çözümlerin değerlendirilmeleri gerekmektedir. Diyabet hastalarının hayatlarını kolaylaştıracağına ve iyileşme süreçlerine yardımcı olacağına inanılan hizmetler yaygın olarak kan şekeri ve tansiyon bilgilerinin tutulması, egzersiz takibi, su tüketim takibi, hatırlatıcı veya uyarıcı mesajlar verilmesi şeklinde listelenmektedir. İncelenen çalışmaların sunduğu hizmetler de Tablo 1’de karşılaştırılmıştır. Çalışma [4], [5] ve [9] bulut bilişimin sağlık hizmetlerinde kullanılmasını önermekle beraber mobil teknolojinin kullanımı bu çalışmalarda yer almamaktadır.

Olmayan MBB özelliğine bağlı olarak, bu çalışmaların sunduğu mobil günlük veya takip uygulamaları da bulunmamaktadır. Sağlık hizmetlerinde hastalara yardımcı olmak için MBB kullanılmasını öneren çalışma [7] ve [8] ise diyabet hastalarını özellikle hedef alan bir öneri içermemektedir. Çalışma [12], diyabet hastalarına yönelik bir MBB uygulaması olmasına rağmen, hastaların şeker seviyesi takibi konusunda tıbbi bir çalışmayı oluşturmakta, fakat kullanıcılarına başka hizmetler sunmamaktadır. Çalışma [10] ve [13] ise sadece günlük hizmeti sunan birer yazılım uygulaması önermektedir. Direk olarak diyabet hastalarını hedefleyen MBB kullanımını öneren ve en fazla hizmeti sunan iki uygulama çalışma [11] ve MD olarak görülmektedir. Çalışma [11] günlük, egzersiz takibi ve su tüketimi takibi hizmetlerini sunsa da, MD ilave olarak uyarıcı mesajlar verme ve düzenli hatırlatmalar yapan özellikleriyle çalışma [11]’den daha fazla hizmet sunmuş olmaktadır. Ancak çalışma [11]’in önerilen MD gibi HL7 desteği bulunmamaktadır.

Tablo 1. Uygulama karşılaştırılma tablosu (Application comparison table)

Çalışmalar	Sunulan Hizmetler							
	Diyabet Takip Uygulaması	Mobil uygulama	Bulut altyapısı	Günlük	Egzersiz	Su Tüketimi	HL7	Uyarıcı Mesajlar(Su ve Egzersiz Hatırlatıcı)
[4]	X	X	√	X	X	X	X	X
[5]	X	X	√	X	X	X	√	X
[7]	X	√	√	X	X	X	X	X
[8]	X	√	√	X	X	X	X	X
[9]	X	X	√	X	X	X	X	X
[10]	√	√	√	√	X	X	X	X
[11]	√	√	√	√	√	√	X	X
[12]	√	√	√	X	X	X	X	X
[13]	√	√	√	√	X	X	X	X
MD	√	√	√	√	√	√	√	√

MD kullanan bir diyabet hastası kan şekeri değerinin yükselme eğilimine girdiği, su tüketimi veya egzersiz düzeyinin yeterli düzeyde olmadığı hakkında uyarılar almaktadır. Ayrıca, MD kullanıcıları geçmiş faaliyetlerini düzenli olarak gerçekleştirdiler de, faaliyet programlarındaki saatlere uygun olarak faaliyetleriyle ilgili hatırlatmalar almaya devam etmektedirler. Bu sayede MD kullanıcıları çok öneme sahip olan günlük ihtiyaçları hakkında sürekli mesajlar almaktadır. MD sahip olduğu özelliklerle diyabet hastalarına sürekli bir yardımcı görevi gören bir şahıs gibi, günün her anında arkadaşlık etmektedir.

5. SONUÇ (CONCLUSION)

Günümüzde akıllı telefon ve mobil cihazlar her yaşta diyabet hastası kullanıcı tarafından yoğun olarak kullanılmaktadır. Çalışmamızda diyabet hastası olanların hayatlarını kolaylaştırmak ve iyileşme süreçlerine yardımcı olmak amacıyla kullanıcı dostu bir yazılım uygulaması önerilmektedir. Önerilen yazılım MBB teknolojisi kullanmakta ve hem sunucu hem de istemci tarafında uygulamalar içermektedir. Önerilen yazılımın diyabet hastalarına sunulan benzer çözümlerdeki uygulamalardan daha fazla hizmet sunan özellikleri bulunmaktadır. Önerilen MD'in birçok özelliği barındırıyor olması, bir diyabet hastasının başka yazılımlara gerek kalmadan çoğu ihtiyacını karşılamasını sağlamaktadır. Benzer çalışmalarla yapılan kıyaslamalar

önerilen yazılımın sahip olduğu özelliklerin diğer çalışmalara nazaran daha fazla olduğunu göstermektedir. Sahip olduğu HL7 özelliğiyle de, MD'nin hem yurtiçi hem de yurtdışındaki diyabet hastalarına yardımcı olabileceği değerlendirilmektedir. Çözümün başka dillere uyarlanması ve doktor gözetiminde kullanıma sunulması istatistiksel çalışmaların yapılması için girişimler devam etmektedir.

KAYNAKLAR (REFERENCES)

- [1]Internet: Türk diyabet cemiyeti, <http://www.diabetcemiyeti.org>,13.05.2016.
- [2]Internet: Diabetesatlas, <http://www.diabetesatlas.org>,13.05.2016.
- [3] J. E. Bibault, P. Blanchard, B. Dubray, E. Lartigau, A. Zadeh, "ELQ: A biologically-equivalent dose calculator available on iPhone, Android, and the web", *Practical Radiation*, 1(3), 212-213, 2011.
- [4] Ms. PrincyMatlani, N. D. Londhe, "A cloud computing based telemedicine service", **IEEE Point-of-Care Healthcare Technologies (PHT)**, Bangalore, India, 326330, 1618 January, 2013.
- [5] O. S. Lupşe, M. M. Vida, L. S. Tivadar, "Cloud computing and interoperability in healthcare information systems", **INTELLI : The First International Conference on Ontelligent Systems and Applications**, Chamonix, France, 8185, 2904 April-May, 2012.
- [6] N. Keskin, H. Kılınc, "Mobil öğrenme uygulamalarına yönelik geliştirme platformlarının karşılaştırılması ve örnek uygulamalar", *Açıköğretim Uygulamaları ve Araştırmaları Dergisi*, 1(3), 68-90, 2015.

- [7] C. K. Yahaya, W. H. Hassa, "Context-Awareness in Mobile Cloud Computing: Healthcare", *The Society of Digital Information and Wireless Communications (SDIWC)*, 30-34, 2013.
- [8] D. B. Hoang, L. Chen, "Mobile cloud for Assistive Healthcare (MoCAsH)", **IEEE Asia-Pacific Services Computing Conference**, Hangzhou, China, 325331, 0610 December, 2010.
- [9] S. J. Weng, L. S. Lai, D. Gotcher, H. H. Wu, Y. Y. Xu, C. W. Yang, "Cloud Image Data Center for Healthcare Network in Taiwan", *Journal of Medical Systems*, 40(4), 1-11, 2016.
- [10] W. C. Hsu, K. H. Lau, R. Huang, H. Le, S. Gilroy, M. Abrahamson, J. Moore, "Utilization of a Cloud-Based Diabetes Management Program for Insulin Initiation and Titration Enables Collaborative Decision Making Between Healthcare Providers and Patients", *Diabetes Technology & Therapeutics*, 18(2), 59-67, 2016.
- [11] E. Arsand, D. H. Froisland, S. O. Skrovseth, T. Chomutare, N. Tatara, G. Hartvigsen, J. T. Tufano, "Mobile health applications to assist patients with diabetes: lessons learned and design implications", *Journal of Diabetes Science and Technology*, 6(5), 1197-1206, 2012.
- [12] C. Quinn, M. Shardell, M. Terrin, E. Barr, S. Ballew, A. Gruber-Baldini, "Cluster-randomized trial of a mobile phone personalized behavioral intervention for blood glucose control", *Diabetes care*, 36(11), 1934-1942, 2013.
- [13] C. C. Quinn, S. S. Clough, J. M. Minor, D. Lender, M. C. Okafor, A. Gruber-Baldini, "WellDoc TM mobile diabetes management Randomized controlled trial: Change in clinical and behavioral outcomes and patient and physician satisfaction", *Diabetes Technology & Therapeutics*, 10(3), 160-168, 2008.

Veri Analizinde İstatistik mi Veri Madenciliği mi?

İlkim Ecem EMRE, Çiğdem SELÇUKCAN EROL

Enformatik Bölümü, İstanbul Üniversitesi, İstanbul, Türkiye

ecememre@gmail.com, cigdems@istanbul.edu.tr

(Geliş/Received:01.08.2016; Kabul/Accepted:20.04.2017)

DOI: 10.17671/gazibtd.309297

Özet— Günümüzün teknolojik gelişmeleri, veri boyutlarının ve çeşitliliğinin artması klasik istatistiksel yöntemlerin yetersiz kalabileceği durumları da beraberinde getirmektedir. Veri analizinde yıllardır kullanılan istatistik, son yıllarda veri madenciliği ile yan yana yer almaktadır. Veri analizi çalışmalarında kullanılan istatistik ve veri madenciliği yöntemleri birçok farklı araştırma alanında kendilerine yer bulabilmektedir. Özellikle sağlık alanında veri analizinde istatistik sıklıkla kullanılmakta, veri madenciliği yöntemleri ise son yıllarda araştırmacılar tarafından fark edilmektedir.

Bu çalışmanın amacı, her iki alanın incelenerek aralarındaki farklılıkların ve benzerliklerin ortaya konmasıdır. Özellikle sağlık alanında veri madenciliğinin çok yaygın olarak bilinmemesi ve sağlık çalışanlarının genellikle istatistiksel yöntemlere bağlı kalıyor olması çalışmanın çıkış noktasını oluşturmaktadır. Çalışma kapsamında veri madenciliği ve istatistiksel yöntemlerle ilgili olarak bir literatür araştırması yapılmıştır. Sonuç olarak bu iki alan kavramsal açıdan incelenmiş, aralarındaki farklar ve benzerlikler özetlenerek ortaya konmaya çalışılmıştır. İlerleyen zamanlarda çalışmanın sağlık verisi ile uygulaması yapılarak genişletilmesi umulmaktadır.

Anahtar Kelimeler— Veri Madenciliği, İstatistik, Veri Analizi

Statistics or Data Mining for Data Analysis

Abstract— Today's technological advances, increase of volume and diversity of data, brings inadequate status of classical statistical methods with it. Statistics, which is being used for data analysis for many years, is located side by side with data mining in recent. Statistics and data mining methods can find place in many different research fields. Especially in health sector statistics is being used frequently, whereas data mining methods is started to get attention of researchers in recent years.

The aim of this study is to observe similarities and differences between two fields. Starting point of the study is that while statistical methods are being frequently used in health field, data mining methods is not known widespread in the same field. In this study, literature research is made about data mining and statistical methods. As a result of the research both fields are observed conceptually, differences and similarities between fields are tried to be summarized in the study. In future it is planned to expand the scope this study with application of data mining and statistical methods to the same health data.

Keywords— Data Mining, Statistics, Data Analysis

1. GİRİŞ (INTRODUCTION)

Günümüzde teknolojinin gelişimi ve veri boyutlarındaki hızlı artış farklı kavramlarında hayatımıza girmesine neden olmuştur. Bilgi iletişim teknolojilerinin hayatın hemen her alanına girmesi, hızlı gerçekleşen teknolojik gelişmeler veri boyutlarındaki türlerindeki artışı tetiklemektedir. Bu nedenle veri yığınları arasında anlamlı, işe yarar bilgi elde edilebilmesi açısından veri analizi her geçen gün daha fazla önem kazanmaktadır. Veri analizinde, veriyi anlamlandırma ve kanıtla dayandırma bilimi olan istatistik, araştırmacıları sınırlandırabilmekte ve büyük boyutlu veri kümeleri karşısında yetersiz kalabilmektedir. İstatistiksel yöntemler birçok alanda kullanılsa da veri analizinde temelini istatistiksel yöntemlerden alan veri madenciliği kavramı

ortaya çıkarak, gerek yapısal gerek yapısal olmayan farklı tipte ve büyük boyuttaki verinin analiz edilebilmesine olanak sağlamaktadır. Veri madenciliği ve istatistik birbirine benzer hedeflere hizmet etse de aralarında bazı farklılıklar bulunmaktadır. Bu farklılıklar ve benzerlikler bu çalışma kapsamında incelenmiştir.

2. KAVRAMLAR (CONCEPTS)

2.1. İstatistik (Statistics)

Bilgiyi kanıtla dayandırma ihtiyacından ortaya çıkan istatistik [1] yüzyıllardır veri analizinde ve veriyi anlamlandırmada kullanılmaktadır. Kökeni veri madenciliğinden çok daha eskilere dayanan istatistik "veriden öğrenme, belirsizliği ölçme, kontrol etme

bilimidir" [2]. Değişkenler arasındaki ilişkileri inceleyen, veri setini çeşitli hesaplamalarla analiz edip bulguları özetleyen, örneklemden çıkartılan sonuçların veri kütlesi için genelleştirmeye yarayan bir bilim dalı olarak istatistik veri madenciliğinden çok daha eski yıllarda ortaya çıkmış ve veri analiz etmede yüzyıllar boyunca kullanılmıştır. İstatistiğin tarihine bakıldığında 16-17. yüzyıllara kadar dayanan çalışmalar olduğu görülmektedir [3].

Ortalama, standart sapma gibi hesaplamalar, değişkenler arasındaki ilişkinin ortaya konması, sonuçların özetlenmesi, geleceğe yönelik tahminde bulunulması gibi çeşitli analizler için istatistikten biliminden yararlanır. Buna bağlı olarak verinin olduğu her alanda veriyi analiz etmek için istatistikten yararlanılabileceğini söylemek yanlış olmayacaktır. Veri madenciliğinin kullanıldığı alanların hemen hepsinde istatistiksel yöntemler de kullanılabilir. Kullanılacak alandan ziyade veri boyutu ve yapılmak istenen analizin niteliği hangi yöntemlerin kullanılacağı konusunda belirleyicidir. Önemli olan bu noktada veri ile ilgili nasıl bir analiz yapılacağına ve veri madenciliği veya istatistik yöntemlerinden hangilerinin kullanılacağına karar verilmesidir.

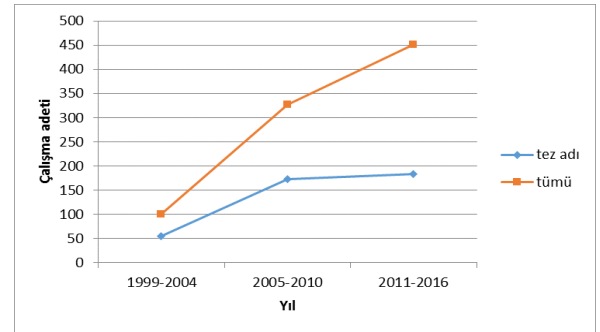
2.2. Veri Madenciliği (Data Mining)

Veri madenciliği veri miktarındaki artışla beraber çokça ilgi görmeye başlayan kavramlardan biridir. Toplamda 9000'den fazla kez alıntılanan Fayyad ve diğ. [4, 5] tarafından yazılan "From Data Mining to Knowledge Discovery in Databases (KDD)" ve "The KDD Process For Extracting Useful Knowledge From Volumes of Data" makaleleri veri madenciliği ve veri tabanlarında bilgi keşfi konularıyla ilgili olarak 90'lı yıllarda oluşturulan en temel kaynaklardır [4], [5]. Her iki makalede de hızla artan veri boyutlarından anlamlı bilgi elde edebilmek için, bilgisayar temelli yeni yöntemlere ihtiyaç duyulduğundan, veri tabanlarında bilgi keşfi ve veri madenciliği konularının araştırmacıların dikkatini çektiğinden bahsedilmektedir [4], [5]. Makalelerin yazılmış olduğu yılın üzerinden 20 yıl geçmesine rağmen bu konular hala daha güncelliğini korumakla beraber bugünün araştırmacıları için büyük öneme sahiptir. Farklı alanlardaki araştırmacılar açısından, veri madenciliğinin bilinirliği ve kullanımı 1996'dan bu yana gözle görülür şekilde artmış ve her geçen gün artmaya devam etmektedir.

Fayyad ve diğ. [5] bilgisayar tabanlı veri madenciliği analizlerin konuşulmaya başlandığı 1960'lardan beri kötü bir üne sahip olduğunu belirtmiştir. Bir kişi eğer yeterince uzun süre veri setini ele alırsa içerisindeki örüntüleri bulabileceği düşünülmektedir. Ancak bu görüşün 1996'da ortaya konduğu göz önünde bulundurulursa o günden bugüne veri madenciliği ile ilgili düşüncelerin değişmiş olduğu ve alanla ilgili bu tip ön yargıların kırılmış olduğu

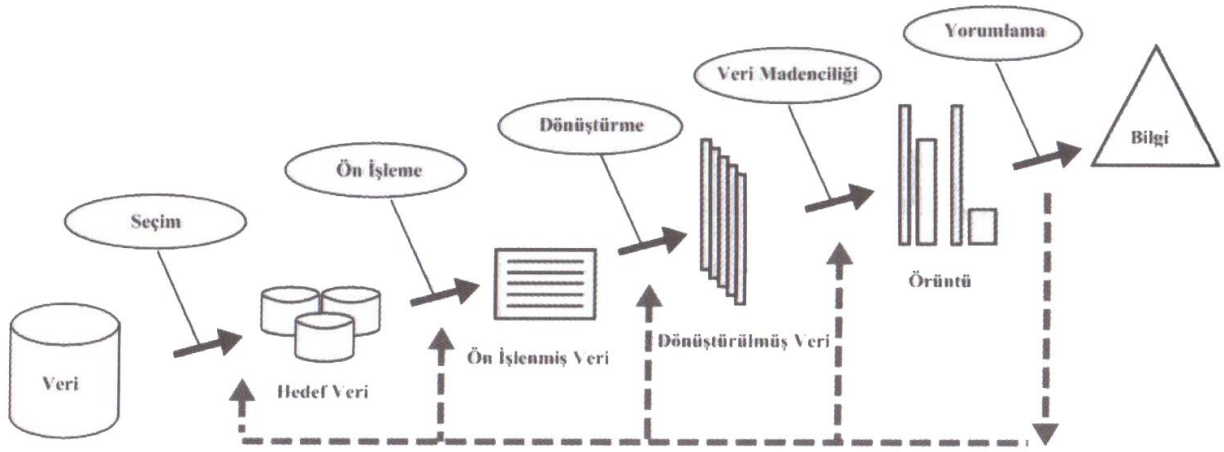
görülmektedir. Ülkemizde yapılan çalışmalara bakıldığında veri madenciliği ile ilgili birçok yayın olduğu görülmektedir. Savaş ve diğ. [6] yaptıkları literatür çalışmasında Türkiye'de veri madenciliği alanında yapılmış çalışmalara yer vermişlerdir. Çalışma kapsamında 2003-2012 yılları arasında eğitim, mühendislik, bankacılık, tıp, ticaret ve telekomünikasyon alanlarında yapılmış olan ve Türkiye'de yayımlanan 43 adet çalışmaya ait bilgiler verilmiştir.

Veri madenciliği ile ilgili çalışmaların yıllar içerisinde arttığını göstermek adına ayrıca Yükseköğretim Kurulu Başkanlığı'na ait Tez Merkezi web sitesinde de bir tarama yapılmıştır [7]. "Veri madenciliği" anahtar kelimeleri iki farklı filtre ile aranmıştır. İlk aramada veri madenciliği tez adları içerisinde, ikinci taramada ise aynı kavram herhangi bir kriter belirtmeden tüm alanlar içerisinde aranmıştır. Şekil 1'de arama sonuçları ve buna bağlı olarak veri madenciliği ile ilgili lisansüstü çalışmaların sayısı verilmiştir.



Şekil 1. Yıllara göre veri madenciliği lisansüstü çalışmaları
(Thesis studies about data mining by years)

Bilgi ve iletişim teknolojilerindeki gelişmeler çok hızlı bir şekilde gerçekleşmekte ve bu durum farklı kaynaklardan elde edilen farklı türlerde (yapısal, yarı yapısal ve yapısal olmayan) verinin birikmesine sebep olmaktadır. Bu veri, günümüzde büyük veri ile ifade edilen bir hale gelmiştir. Verinin hızlı artışı insanların bu veri yığınlarını yorumlayabilmelerini zorlaştırmaktadır. Bu yüzden büyük hacimli verinin analizi ya da yorumlanması için araştırmacılar veri madenciliği alanına yönelmektedir [8].



Şekil 2. Veri tabanlarında bilgi keşfi [5, 10]
(Knowledge discovery in databases)

Veri tabanlarında bilgi keşfi (*Knowledge Discovery in Databases - KDD*) kavramı adından da anlaşılacağı gibi büyük veri tabanlarında aşamalı bir şekilde bilgi keşfedilmesi sürecini ifade eder. Veri madenciliği ise bu keşif sürecinin belki de en önemli / kilit rol oynayan aşamasıdır [5]. Bir çok farklı alandan beslenen veri madenciliği; istatistik, makine öğrenmesi, yapay zeka, örüntü tanıma alanlarındaki gelişmelerden de beslenmektedir [9]. Akpınar'a [10] göre veri madenciliği yöntemleri istatistik, yapay zeka ve makine öğrenmesinden yararlanmaktadır.

Fayyad ve diğ. [5] veri tabanlarında bilgi keşfi sürecini "verideki geçerli, özgün, potansiyel olarak kullanışlı ve anlaşılabilir desenleri belirlemeye yarayan önemli bir süreç olarak" tanımlamıştır (Şekil 2). Veri madenciliği ile ilgili çalışan araştırmacılar ise bu tanımlamayı veri madenciliği kavramını tanımlamak için kullanmışlardır. Bu noktada KDD ve veri madenciliğinin tanımının birbirleri yerine kullanılabilirliği görülmektedir. Kuonen [1], Kumar ve Bhardwaj [9] ve Kapoor [11] çalışmalarında; Fayyad ve diğ. [5]'nin KDD için yaptığı tanımlama veri madenciliğini tanımlamak için kullanmıştır. Veri madenciliği, veri tabanlarında bilgi keşfi sürecinin bir parçası olmasına rağmen tüm sürecin anlamlı sonuçlara varması açısından belki de en önemli aşamasıdır. Elbette diğer süreçlerin analizler açısından önemi göz ardı edilemez ancak KDD içinde en fazla öne çıkan aşama veri madenciliğidir. Bu sebepten dolayı KDD için Fayyad ve diğ. [5] tarafından yapılan tanımlamanın, zaman içinde farklı araştırmacılar tarafından veri madenciliğinin tanımlanması için kullanmaya başlanması normal karşılanmaktadır.

Veri madenciliğinin genel bir tanımı yapılacak olursa, büyük hacim ve farklı tipteki veriden anlamlı bir takım desenler/örüntüler oluşturmayı, veri arasındaki ilişkileri keşfetmeyi ve veriden bilgi elde etmeyi amaçlayan bir alandır. Tıpkı bir maden keşfi yapar gibi veri yığınları arasından da "değerli, anlamlı" olan bilginin bulunmasını sağlar.

Özkan [12] veri madenciliğini "Büyük ölçekli veriler arasından değeri olan bir bilgiyi elde etme işidir." şeklinde tanımlamıştır.

Genel geçer bir tanım olmamakla beraber veri madenciliği için birbirine benzer birçok tanım yapılmaktadır:

"Veri madenciliği büyük hacimli veriler arasından bilgiyi çekip çıkarmaktır. Başka bir deyişle veri madenciliği, veri tabanlarındaki büyük ve karmaşık veriler arasındaki bilinmeyen örüntüleri, değerli yapıları ve ilginç ilişkilerin keşfedilmesi bilimidir." [13]

"Veri madenciliğindeki esas amaç veri setinden bilgi elde etmek ve bunu anlaşılabilir bir yapı halinde kullanılmak üzere ortaya koymaktır." [11]

"Veri madenciliğinin amacı, geçmiş faaliyetlerin analizini temel alarak gelecekteki davranışların tahminine yönelik karar-verme modelleri yaratmaktır." [14]

Özetlemek gerekirse farklı kaynaklarda yapılan farklı tanımlar aslında birbirine çok yakın noktalara parmak basmaktadırlar. Veri madenciliği de istatistik gibi çeşitli analizler yoluyla veriden öğrenmeye ve bilgi elde etmeye yarayan bir alandır. Esas olarak büyük boyutlu veri tabanlarındaki karmaşık yapıları verinin arasındaki örüntülerin bulunmasını, anlamlı ve kullanışlı olacak bilgilerin keşfedilmesini, elde edilen bilginin anlaşılabilir bir şekilde ortaya konmasını sağlar. Bunun yanında veri madenciliği ile istatistiksel analizlerde olduğu gibi geleceğe dönük tahminler yapmak da mümkündür [1], [4], [5], [8], [12], [13], [15], [16].

Veri madenciliği birçok farklı araştırma alanında kullanılmaktadır (Tablo 1).

Tablo 1. Veri madenciliğinin kullanım alanları
(Data mining application areas)

Benzerlikler	Detaylar	
	Uygulama	Kaynak
Sağlık / Tıp	Hastalıkların tahmini ve analizi Hastalıkların etkilerinin analizi İlaç yan etkilerinin tespiti Hastane yönetimi	[14], [17], [18]
Biyoenformatik	Biyolojik veri analizi	[19]
Astronomi	Gök cisimlerinin analizi/sınıflandırılması	[4]
	İklim modellemeleri	[9]
Telekomünikasyon	Müşteri kayıp analizi Müşteriye özel kampanya analizi	[6]
Üretim	Hata tespiti ve tahmini	[4]
Yatırım	Portfolyo yönetimi	[4]
Pazarlama	Müşteri davranış tahminleri Uygun kampanya geliştirme Farklı müşteri gruplarının analizi Pazar sepet analizi	[4], [15]
Bankacılık	Dolandırıcılık tespiti Para aklama işlemlerinin tespiti Müşteri gruplarının belirlenmesi	[4], [15]
Sigortacılık	Dolandırıcılık tespiti Müşteri tahmini	[15]
Yapısal olmayan veri analizi	Sosyal medya verisi Web sitelerinin analizi	[20]
Eğitim	Başarı faktörlerinin belirlenmesi	[6], [21]

3. YÖNTEM (METHOD)

Bu çalışmanın oluşturulması için literatür taraması yapılmıştır. 1996 ve 2015 yılları arasında yayımlanan konferans bildirisi, makale ve kitap olmak üzere, Türkçe, İngilizce ve Almanca dillerinde yazılmış toplam 24 kaynak incelenmiştir. İnternet üzerinden Google arama motorunda yapılan taramalarda "veri madenciliği", "data mining", "istatistik", "veri madenciliği ve istatistik", "veri madenciliği ve istatistiğin farkları", "differences between data mining and statistics", "unterschiede zwischen statistik und data mining" anahtar kelimeleri kullanılmış ve çıkan sonuçlar taranarak aralarından çalışma için uygun bulunan 24 kaynak incelenmiştir.

4. BULGULAR (FINDINGS)

4.1. İstatistik ve Veri Madenciliği Arasındaki Benzerlikler (Similarities Between Data Mining and Statistics)

İstatistik ve veri madenciliği kavramsal olarak farklı alanlar olsa da ikisinin ortak özellikleri vardır (Tablo 2). Her iki alan da veriden öğrenme, verinin bilgiye dönüştürülmesi, veriyi analiz etme, verinin anlamını çözme, belirsizlikleri ortadan kaldırma, olayı etkileyen faktörleri belirleme, ön görüde bulunma amaçlarını yerine getirirler [1], [13], [16], [22]. Genel amaçlarına bakıldığında her iki alanın da odaklandığı nokta veriden öğrenme ya da veriyi bilgiyi dönüştürmedir. Bu açıdan, her ikisinin de edindiği misyon birbirine çok yakındır.

Tablo 2. İstatistik ve veri madenciliğinin benzerlikleri
(Similarities between data mining and statistics)

Benzerlikler
Veriden öğrenme / verinin bilgiye dönüştürülmesi
Veriyi analiz etme / verinin anlamını çözme
Öngörüde bulunma / tahmin yapma
Belirsizlikleri ortadan kaldırma
Olayı etkileyen faktörleri belirleme
Veri ön işleme
Kullanılan analiz çeşitleri

[1], [9], [13], [15], [16], [22]

Veri madenciliği yöntemlerinin birçoğu temelinde istatistiksel yöntemlere dayanır. Başka bir deyişle veri madenciliğinin temelini istatistik oluşturur ve istatistik bilimi olmadan veri madenciliğinden söz etmek mümkün olamaz [9]. Bazı araştırmacılara göre istatistiksel bakış açısıyla bakıldığında veri madenciliği yöntemleri daha esnek yöntemler olarak tanımlanmış ve veri madenciliği istatistikteki çok değişkenli analizlerle eş değer tutulmuştur [1].

Kullanılan analizler incelendiğinde; kümeleme, diskriminant, regresyon, korelasyon analizlerinin her iki alanda da kullanılan ortak yöntemler olduğu görülmektedir [16]. Aslında veri madenciliği istatistikte var olan bu analizleri kullanmakta ve bu açıdan bakıldığında veri madenciliği, yukarıda da bahsedildiği gibi çok değişkenli istatistiksel analizler olarak değerlendirilebilmektedir. Yine veri madenciliğinde sıklıkla kullanılan karar ağaçları veya birliktelik analizleri de çok değişkenli istatistik analizleri ile eş tutulmaktadır [13].

Her iki alan için kullanılan yöntemler dışında sahip olunan bir diğer benzerlik ise veri ön işleme aşamasıdır. Ön işleme aşaması, verinin analize girmeden önce işlenmeye uygun hale getirilmesini sağlarken, analizin doğruluğu ve sonuçların kalitesi için de önemli rol oynamaktadır. Analiz sonuçlarının doğruluğu için analize giren verinin de temiz olması gerekir [16]. Akpınar [15] ön işlem aşamasının veri tabanlarında bilgi keşfi sürecinin %50 ile %80' ini oluşturabileceğini belirtmiştir.

4.2. İstatistik ve Veri Madenciliği Arasındaki Farklılıklar (Differences Between Data Mining and Statistics)

İstatistik ve veri madenciliği arasında esas amaçları birbirine çok yakın olsa da önemli farklılıklar olduğu görülmektedir (Tablo 3).

Tablo 3. İstatistik ve veri madenciliğinin farklılıkları
(Differences between data mining and statistics)

Farklılıklar	Detaylar	
	İstatistik	Veri Madenciliği
Kavramsal [22]	Kökene eskiye dayanan, başlı başına bir bilim dalıdır	İstatistiğin alt dalı değildir.
Veri seti büyüklüğü [12], [13], [16]	Görece daha küçük boyutlu veri	Milyonlarca, milyarlarca veri, çok fazla değişken
Örneklem büyüklüğü [16]	Veri setinden seçilen bir küme	Veri setinin tamamı
Verinin toplama amacı [13], [16]	Belli bir amaç için toplanır	Birincil amaç veri madenciliği uygulamak değildir. (genellikle)
Hipotezin varlığı [16], [23], [24]	Hipotez var.	Hipotez yok/olmayabilir.
Benimsenen yaklaşım [16]	Tümevarım	Tümdengelim
Bilgisayar kullanımı [22]	Bilgisayarsız analiz mümkündür.	Bilgisayarsız düşünülemez.

Kavramsal açıdan her iki kavramın birbirinden bağımsız olduğunu bilmek önemlidir. Veri madenciliği ve istatistik birbirlerinden farklı alanlardır ve veri madenciliği istatistiğin bir alt dalı ya da parçası değildir. Temelinde istatistiğe dayanan birçok veri madenciliği yöntemi / algoritması vardır ancak bunlar sebebiyle veri madenciliğini istatistiğin bir alt dalı olarak görmek yanlıştır [22].

Analizlerde kullanılan veri setlerinin büyüklüğü açısından bakıldığında da her iki alan farklılık göstermektedir.

Geleneksel olarak kullanılan istatistiksel yöntemler çok büyük hacimli veri setleri, büyük veri gibi kavramlar ile karşılaştığında, verinin büyüklüğü ve çeşitliliği açısından yetersiz kalabilmektedir. Bu sebeple veri madenciliği yöntemleri öne çıkmakta ve büyük veri setlerinin analizinde kullanılmaktadır. Veri madenciliği yöntemleri için kolaylıkla analiz edilebilecek veri setleri istatistiksel yöntemlerle analiz edilemeyecek büyüklükte olabilir [13].

"Bir istatistikçi için büyük veri kümesi birkaç yüz veya bin veri kümesi içerir. Veri madenciliği ile uğraşanlar için, milyon veya milyar veri beklenmeyen bir durum değildir." [16]

"Veri madenciliğinde ise milyonlarca ve hatta milyarlarca veri ve çok fazla değişken ile ilgilenilir." [12]

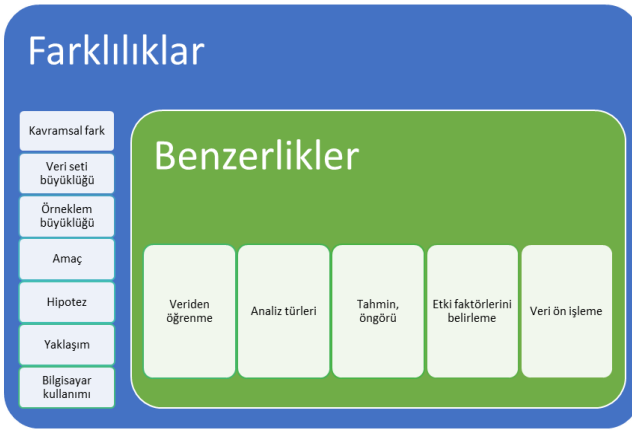
Veri setinin büyüklüğü gibi örneklemin büyüklüğü de farklıdır [16]. Örneklem, istatistiksel analizlerde ana kütle temsil eden daha ufak bir küledir. Veri madenciliğinde ise örneklem veri setinden seçilen daha ufak bir veri seti değil çoğunlukla eldeki veri setinin tamamıdır. Bu sebeple veri madenciliği analizleri bilgisayar kullanımı olmadan düşünülemez [22]. Çünkü eldeki veri yığını farklı tipte veri içermekle beraber çok fazla sayıda veriyi içerir ve veri madenciliği algoritmalarını, karmaşık ve büyük hacimli veri yığınlarına bilgisayar olmadan uygulamak neredeyse imkansızdır. Oysa istatistiksel analizlerde veri seti veri madenciliğindeki orana daha küçük olduğundan bilgisayar kullanmadan da analizler yürütülebilir. Bu duruma başka bir açıdan bakılacak olursa istatistiğin kökenleri veri madenciliğinden ve bilgisayarların gelişiminden daha eskilere dayandığından, zaten istatistiksel analizler uzunca bir süre bilgisayar kullanımı olmadan yapılmıştır [22].

Analizi yapılacak verinin toplama biçimi veya amacı açısından bakıldığında da veri madenciliği ve istatistiğin yaklaşımı farklılık göstermektedir. İstatistik çalışmaları kapsamındaki veri belirli bir amaç için veya sorudan yola çıkılarak toplanır [13], [16]. Bu duruma bağlı olarak başka bir fark da göz önüne alınmalıdır; o da hipotezin varlığıdır.

Önceden belirlenmiş bir soruyla yola çıkılan istatistiksel araştırmalarda, analiz ve sonuç aşamalarından önce kurulan bir hipotez mevcuttur. Veri madenciliğinde ise analizden önce tanımlanmış bir hipotezin varlığından söz edilemez [23]. Hipotezden yola çıktığı için istatistiksel analizlerde apriori bilgiden yola çıkılarak analize başlandığı söylenebilir. Apriori bilgi deneyden veya herhangi bir deneyimden bağımsız olarak var olan bilgi olarak adlandırılabilir [25]. Bu şekilde yola çıkılarak analiz sonucunda ortaya konan hipotezin doğruluğu veya yanlışlığı ispatlanmış olur. Veri madenciliği analizlerinin ise böyle bir iddiası yoktur. Çünkü analize başlamadan önce herhangi bir hipotez belirlenmez. Aksine analiz sonucunda elde edilen bulgular incelenir ve yorumlanır. Veri madenciliği ve istatistikte bu sebeplerden ötürü hipotezin rolü farklılık gösterir [16], [24].

İki alan arasındaki diğer fark ise analiz yapılırken benimsenen yaklaşımdır. Tümevarım ya da tümdengelim yaklaşımlarından hangisinin benimsendiği farklı kaynaklarda farklı şekillerde açıklanmıştır. Tüzüntürk [16], veri madenciliği yöntemlerinin kullanıldığı analizlerde tümdengelim, istatistiksel analizlerde tümevarım yaklaşımlarının benimsendiğinden bahsetmektedir. Veri madenciliği analizlerinde veri setinin tamamının analiz edilmesi ile daha özel/yerel bilgiye ulaşılır. Bu durumda genelden özele doğru bir bilgi keşfi olduğu yani tümdengelim yaklaşımıyla ilerlendiği söylenmektedir. İstatistiksel analizlerde ise tam tersine, tümevarım yaklaşımı ile hareket edilir çünkü ana kütle içinden seçilen örneklemin analizinden yola çıkılarak elde edilen sonuçlar tüm ana kütle/veri seti için genellenir. Bu yaklaşımda özel olandan genel olana doğru bir yönelme vardır. Ganesh [13] ise veri madenciliğinin genellikle asıl toplanma amacından bağımsız olan ikincil analizler yapmada kullanıldığını belirtmiştir. Bu sebeple tümevarım yaklaşımıyla hareket edildiğini söylemektedir. İstatistiksel analizler ise ona göre hipotezden yola çıkılarak tümdengelim yöntemiyle ilerlemektedir. Bu çalışmada kapsamında; veri madenciliğinin tümdengelim, istatistiksel yöntemlerin tümevarım yöntemlerine uygun olarak yapıldığı görüşü hakim olmaktadır.

Çalışma kapsamında elde edilen benzerlik ve farklılık Şekil 3'te özetlenerek sunulmuştur:



Şekil 3. Farklılıklar ve benzerlikler
(Differences and similarities)

5. TARTIŞMA VE SONUÇ (DISCUSSION AND CONCLUSION)

Günümüzdeki teknolojik gelişmeler ve veri boyutlarındaki artış, istatistiksel yöntemlerin tek başına yetersiz kalabilmesine sebep olmaktadır. Bu gelişmeler sonucu ortaya çıkan veri madenciliği kavramı büyük veri boyutlarında analiz yapılmasını sağlamaktadır. Bu çalışma kapsamında veri madenciliği ve istatistik alanlarının birbirleriyle nasıl bir ilişki içerisinde olduğu ortaya konmaya çalışılmıştır. Çalışmanın, her iki alan arasındaki ilişki, benzerlik ve farklılıklar hakkında genel

bir çerçeve çizerek araştırmacılara yardımcı olması umulmaktadır.

Veri madenciliği kavramı ayrı bir alan olsa da temeli istatistiksel yöntemlere dayanmaktadır. Veriden öğrenme, anlamlı bilgi elde etme, öz bilgi keşfi, bilgi keşfi, örüntüleri keşfetme, değerli bilgiye ulaşma, değişkenler arasındaki ilişkileri keşfetme gibi farklı şekillerde adlandırılabilir ve analizlere dayalı işlemler, istatistik veya veri madenciliği yöntemleri kullanarak yapılabilmektedir. Bu noktada eldeki veri ile nasıl bir analiz yapılacağına belirlenmesinde analizi yapacak kişinin belirlemesi gerekecektir. Hangi yöntemlerin kullanılacağına belirlerken veri kümesinin büyüklüğünü ve yapısını bilmek; analiz yöntemleri arasındaki farklılıklara hakim olmak önemlidir. Veri madenciliği ve istatistik arasında farklılıklar olmasına rağmen iki alan birbirlerinden bağımsız düşünülemez alanlardır. Çünkü her iki alan da "veriden öğrenme" ve "veriden bilgi elde etme" amaçları için kullanılırlar. Veri madenciliğindeki gelişmelerin istatistiksel yöntemlerin kullanımına, istatistiksel yöntemlerdeki gelişmelerin ise veri madenciliğinin gelişimine katkı sağlayacağı ve bu iki alanın birbirine doğru bir büyüme içinde olduğu düşünülmektedir [1].

İstatistiksel analizler halen veri analizinde kullanılmakta olan analiz yöntemleridir. Ancak teknolojinin gelişmesi, çok çeşitli veri kaynaklarının oluşması, veri üretiminin artması ve veri türlerinin çeşitliliği analiz yöntemlerinin de gelişmesini ve çeşitlenmesini sağlamaktadır. Buna paralel olarak araştırma alanları arasındaki keskin sınırlar ortadan kalkmakta ve her farklı alan birbirinden beslenmektedir. Veri madenciliğinin de beslendiği temel alanlardan biri istatistiktir. Temellerini istatistikten alan veri madenciliği yöntemleri analizlerde istatistiksel yöntemlerin sınırlılıklarını ortadan kaldırmakta ve günümüzün veri bolluğunda her alandan araştırmacılara daha geniş kapsamlı ve esnek analizler yapma imkanı sunmaktadır. "Günün gereksinimlerine göre kendini güncelleyerek gelişimini sürdüren veri madenciliği" [10] her geçen gün farklı araştırma alanlarında araştırmacılar tarafından ilgi görmektedir. Tıp gibi, araştırmacıların istatistiksel yöntemlere bağlı kaldığı alanlarda ilerleyen zamanlarda veri madenciliği yöntemlerinin daha fazla ilgi göreceği düşünülmektedir. Elektronik ortama geçildikten sonraki süreç düşünüldüğünde veri tabanlarında her geçen gün hastalara ait daha fazla veri birikmektedir. Bu verinin analiz edilmesi yoluyla, hastalıklar hakkında öngörülebilir bulunulabilmesi, çeşitli belirtilere ya da özelliklere göre hastalıkların tahmin edilmesi, hastalıklara yönelik önleyici tedbirlerin alınabilmesi gibi araştırma başlıklarında veri madenciliği yöntemlerinin ileride kullanılacağı öngörülmektedir.

Bu çalışmanın ilerleyen aşamalarında aynı veri seti üzerinde hem istatistiksel yöntemlerin hem de veri madenciliği yöntemlerinin uygulamasının yapılması ve sonuçların/yöntemlerin karşılaştırılması planlanmaktadır.

KAYNAKLAR (REFERENCES)

- [1] D. Kuonen, "Data Mining and Statistics: What is the Connection?," The Data Administration Newsletter, 2004.
- [2] M. Davidian and T. A. Louis, "Why Statistics?," Science, 336, 12, 2012.
- [3] İnternet: J. Aldrich, "Figures from the History of Probability & Statistics," Ekim 2012. <http://www.economics.soton.ac.uk/staff/aldrich/Figures.htm>, 31.05.2016.
- [4] U. Fayyad, G. Piatetsky-Shapiro, and P. Smyth, "From data mining to knowledge discovery in databases," AI Mag., 17 (2), 37, 1996.
- [5] U. Fayyad, Piatetsky-Shapiro, and P. Smyth, "The KDD Process for Extracting Useful Knowledge from Volumes of Data," Communications Of The ACM, 39(11), 27–34, 1996.
- [6] S. Savaş, N. Topaloğlu, and M. Yılmaz, "Veri madenciliği ve Türkiye'deki Uygulama Örnekleri," İstanbul. Ticaret Üniversitesi Fen Bilimleri Dergisi, 11(21), 1–23, 2012.
- [7] İnternet: YÖK, "Ulusal Tez Merkezi," 2016. <https://tez.yok.gov.tr/UlusalTezMerkezi/giris.jsp>, 29.06.2016.
- [8] M. E. Balaban and E. Kartal, Veri Madenciliği ve Makine Öğrenmesi, 1. baskı, Çağlayan Kitapevi, İstanbul, 2015.
- [9] D. Kumar and D. Bhardwaj, "Rise of Data Mining: Current and Future Application Areas," International Journal of Computer Science Issues, 8(5), 256–260, 2011.
- [10] H. Akpınar, Data: Veri Madenciliği Veri Analizi, 1.baskı, Papatya Yayıncılık Eğitim, İstanbul, 2014.
- [11] A. Kapoor, "Data Mining: Past, Present and Future Scenario," International Journal of Emerging Trends & Technology in Computer Science (IJETTCS), 3(1), 95–99, 2014.
- [12] Y. Özkan, Veri Madenciliği Yöntemleri, 1.baskı, Papatya Yayıncılık Eğitim, İstanbul, 2008.
- [13] S. Ganesh, "Data mining: Should it be included in the statistics curriculum?," The 6th International Conference on Teaching Statistics (ICOTS 6), Cape Town, Güney Afrika, 2002.
- [14] A. S. Koyuncugil and N. Özgülbaş, "Veri Madenciliği: Tıp ve Sağlık Hizmetlerinde Kullanımı ve Uygulamaları," Bilişim Teknolojileri Dergisi, 2(2), Mayıs 2009.
- [15] H. Akpınar, "Veri Tabanlarında Bilgi Keşfi ve Veri Madenciliği," İÜ İşletme Fakültesi Dergisi, 29(1), 1–22, 2000.
- [16] S. Tüzüntürk, "Veri Madenciliği ve İstatistik," Uludağ Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi, 29(1), 65-90, 2010.
- [17] S. Güllüoğlu, "Tıp ve Sağlık Hizmetlerinde Veri Madenciliği Çalışmaları: Kanser Teşhisine Yönelik Bir Ön Çalışma," AJIT-E AJIT-e: Online Academic Journal of Information Technology, 2(5), 2011.
- [18] E. Kaya, M. Bulun, and A. Arslan, "Tıpta Veri Ambarları Oluşturma ve Veri Madenciliği Uygulamaları," Akademik Bilişim, Çukurova Üniversitesi, Adana, 3-5 Şubat, 2003.
- [19] Ç. Erol and Y. Özkan, "Temel Biyoloji ve Mikrodizi," in Biyoenformatik DNA Mikrodizi Veri Madenciliği, 2.baskı, Papatya Yayıncılık Eğitim, İstanbul, 51–81, 2015.
- [20] M. Ö. Dolgun, T. Güzel Özdemir, and D. Oğuz, "Veri Madenciliğinde Yapısal Olmayan Verinin Analizi: Metin ve Web Madenciliği," İstatistikçiler Dergisi, 2, 48–58, 2009.
- [21] Ç. Kurt and O. A. Erdem, "Öğrenci Başarısını Etkileyen Faktörlerin Veri Madenciliği Yöntemleriyle İncelenmesi," Politeknik. Dergisi., 15(2), 111-116, 2012.
- [22] D. Meschenmoser, "Data Mining und Statistik: Gemeinsamkeiten und Unterschiede," Şubat 2004.
- [23] H. Mannila, "Data Mining: Machine Learning, Statistics and Databases," SSDBM '96 Proceedings of the Eighth International Conference on Scientific and Statistical Database Management, Stockholm, 2–9, 1996.
- [24] C.-M. Zhao and J. L. Vice, "Data mining: Going beyond traditional statistics," New Directions for Institutional Research., 131, 7–16, Eylül 2006.
- [25] İnternet: J. S. Baehr, "A Priori and A Posteriori", <http://www.iep.utm.edu/apriori/>, 22.06.2016.

Scrum Metodu Kullanılarak Bir Mobil Uygulama Geliştirme Sürecinin Gerçekleştirilmesi

Mehmet ELİBOL¹*, Çiğdem SELÇUKCAN EROL²

¹Bilgisayar ve Öğretim Teknolojileri Eğitimi Bölümü, Marmara Üniversitesi, İstanbul, Türkiye

²Enformatik Bölümü, İstanbul Üniversitesi, İstanbul, Türkiye

mehmet.elibol@marmara.edu.tr, cigdemselcukcan@istanbul.edu.tr

(Geliş/Received:23.08.2016; Kabul/Accepted:04.04.2017)

DOI: 10.17671/gazibtd.309299

Özet— Bu çalışmanın amacı, Scrum metodu kullanılarak geliştirilen bir mobil uygulama sürecindeki deneyimlerden yola çıkılarak bu süreçle ilgili önerilerde bulunmaktır. Bu kapsamda klinik araştırmalar alanında ihtiyaç duyulan bir mobil uygulamanın, Scrum metodu ile gerçekleştirilme süreci detaylı bir şekilde anlatılmaktadır. Bu süreçte ön plana çıkan noktalar, projeye başlamak için ekiplerin nasıl oluşturulduğu, müşterinin ihtiyaçlarının nasıl tespit edildiği, süreç içinde değişen müşteri isteklerinin nasıl yönetildiği, bu süreçte şeffaflık, denetim ve adaptasyonun nasıl gerçekleştirildiği Scrum sürecindeki kavramlar (Koşu, Ürün İçeriği, vb.) baz alınarak detaylandırılmıştır.

Anahtar Kelimeler— Mobil Uygulama Geliştirme, Çevik Metodlar, Scrum

Implementation of a Mobile Application Development Process Using Scrum Method

Abstract— The purpose of this study is to suggest the process from the experiences of a mobile application developed using the Scrum method. In this context, the process of carrying out a mobile application, which is needed in the field of clinical researches, with the Scrum method is explained in detail. The concepts in the Scrum process (Sprint, Product Backlog, etc.) are based on the points that are at the forefront of this process, how the teams are set up to start the project, how customer needs are identified, how customer requests are changed in the process and how transparency.

Keywords— Mobile Application Development, Agile Methods, Scrum

1. GİRİŞ (INTRODUCTION)

Günümüzde mobil cihaz kullanımının artması ile birlikte mobil uygulama kullanımı konusunda da artış olmuş ve mobil uygulamalara olan ihtiyaç artmıştır. İhtiyaç duyulan mobil uygulamaların geliştirilmesi sırasında, mobil cihazların cihaz boyutu, ekran boyutu ve veri giriş mekanizması için fiziksel özelliklerinin olmasının yanı sıra işlem gücü, bellek alanı ve işletim sistemi gibi teknik karakterleri de olduğu dikkat edilmesi gereken noktalar [1]. Mobil uygulamaların tasarımı, değişen ihtiyaç ve beklentilerle birlikte [2], yazılım geliştirme projelerinde görünenden çok daha karmaşıktır, bu durum geliştiricileri geleneksel yazılım geliştirme metodolojilerini tekrar gözden geçirmeye zorlamaktadır [3].

Her gün binlerce uygulamanın ortaya çıktığı pazarda ihtiyaç tespit edilerek geliştirilmeye başlayan uygulama bir başka kişi ya da kurum tarafından hazırlanarak hayata geçirilebilir. Uygulama başka firma tarafından hayata geçirildikten sonra uygulamamız kullanılamaz konuma

düşebilmektedir. Bu da uygulama için harcanan zaman ve paranın boşa gitmesi anlamına gelmektedir. Görüldüğü gibi mobil uygulama geliştirme alanında ihtiyacı tespit ederek anında uygulama geliştirmeye başlamak önemli hususlardan birisidir. Bu ihtiyaçları en doğru şekilde tespit edebilmek için müşterinin ne istediğini tam olarak anlamak çok önemlidir. Hızlı gelişen ve büyüyen yazılım dünyasında ihtiyaçları tam olarak tespit edebilmenin bazı zorlukları vardır. Proje başında tespit edilen ihtiyaçlar değişen ve gelişen teknoloji ile birlikte proje sonunda değerini yitirebilmektedir ve ihtiyaç olmaktan çıkmakta hatta gereksiz konuma düşebilmektedir. Bundan dolayı ihtiyaç tespit edildikten sonra müşterinin ihtiyaçlarının ürünün her aşamasında değişebileceği göz önüne alınarak sürekli müşteri ile irtibat halinde bulunmak [4] önem arz eden diğer bir konu olmaktadır. Tüm bu hususlar mobil uygulama geliştirme sürecinde geleneksel yazılım geliştirme metodolojilerinin sınırları olmakta ve bizleri çevik metodolojilere yönlendirmektedir.

2. ÇEVİK YAZILIM GELİŞTİRME METODOLOJİSİ (AGILE SOFTWARE DEVELOPMENT METHODOLOGY)

Çevik Yazılım Geliştirme Manifestosu, projede çalışan bireyler ve aralarındaki etkileşimi ön plana çıkararak, süreçte sürekli müşteri ile işbirliği içinde değişikliklere yanıt verilmesi gerektiğini ve bu süreçte yazılımın sürekli çalışır olması gerektiğini savunan bir yazılım geliştirme metodolojisidir [4].

Çevik metodolojiler, şirketlerin doğru ürün oluşturmaya yardımcı olmak için [2] çevre ve gereksinimlerdeki değişikliklere uyum sağlamayı kolaylaştırmaktadır. Buna ek olarak süreç içindeki geliştirme tecrübelerinden yeni öğrenmeler sağlayarak yazılım geliştirme esnekliğini ve üretkenliğini artırmakta [3] ve şirketlerin pazarda mümkün olduğunca rekabetçi olmasını sağlamaktadır [2]. Örneğin; elektronik imzanın geliştirilmesinden sonra yazılım şirketleri elektronik imza ile işlem yapabilen sistemleri hayata geçirmeye başlamışlardır. Elektronik imza ile çalışan sistemler yaygınlaştıkça elektronik imza ile işlem yapan kişi sayısı da artmaya başlamıştır. Bu süreç devam ederken, teknolojiye gelişmelerle birlikte mobil imza kavramı ortaya çıkmıştır. Mobil imza kavramının ortaya çıkması ile birlikte elektronik imza ile çalışan projelerin revize edilmesi gerekmiştir [5]. Bu noktadan hareketle, görüldüğü gibi yazılım projeleri başka alanlardaki projelere kıyasla değişime daha duyarlıdır. Bu açıdan ele alındığında yazılım projeleri her zaman değişime ve gelişime açık olmalıdır.

3. MOBİL UYGULAMA GELİŞTİRME SÜREÇLERİNDE UYGULANAN ÇEVİK YAKLAŞIMLAR (AGILE APPROACHES FOR MOBILE APPLICATION DEVELOPMENT PROCESS)

Birçok çevik yazılım geliştirme metodu bulunmaktadır. Mobil uygulama geliştirme sürecinde kullanılan çevik yaklaşımlarla ilgili bilgi aşağıda yer almaktadır:

MOBİL-D mobil yazılım ürünleri ve servisleri geliştiren üç firma ile işbirliği içinde geliştirilmiştir. Mobile-D yaklaşımı kısa süre içinde tam fonksiyonlu bir mobil uygulamayı yayına hazır hale getirmek için bir ofis ortamında çalışan on kişiden daha az geliştiriciden oluşan bir ekip için optimize edilmiştir [6].

MASAM bir mobil uygulamayı hızlı şekilde geliştirmek için çevik yazılım geliştirme ilkelerini uygulayan bir yaklaşımdır [7].

Scharff ve Verma [1] Pace Üniversitesinde gerçek bir kullanıcı için mobil uygulama geliştirme sürecinde profesyonel bir Scrum Uzmanı ile birlikte Scrum Metodunu uygulamıştır.

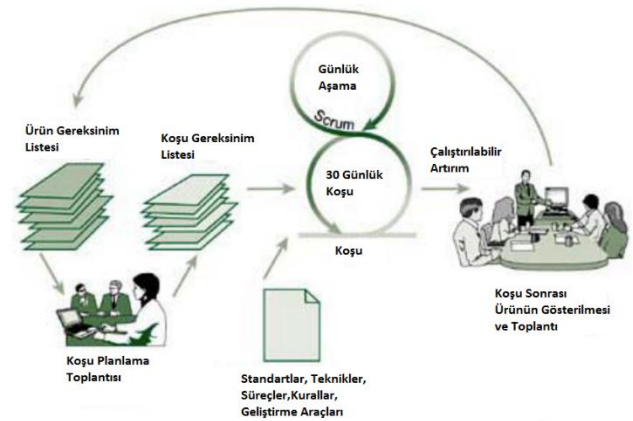
Bu çalışma kapsamında sık kullanılan metodolojilerden biri olan Scrum tercih edildiğinden sadece Scrum çevik yazılım geliştirme metodu ile ilgili bilgi verilmektedir.

3.1. Scrum (Scrum)

Schwaber ve Sutherland (1996) tarafından tanıtılan, hala desteklenen ve Beedle'nin işbirliği ile geliştirilen Scrum karmaşık ürün geliştirmeyi yönetmek için bir süreç çerçevesi sunmaktadır [8, 9]. Scrum'un mobil sistemlere uygulanması ilk defa [1] tarafından bir sınıf ortamında gerçekleştirilmiştir.

Scrum'un merkezinde geliştirme sürecinde ortaya çıkacak süreçlerin tahmin edilemez olduğu fikri yer alır. Bu nedenle, esnek bir şekilde yazılım geliştirme nasıl olabileceği üzerine yoğunlaşmakta [10] ve geliştirme aşaması için herhangi bir özel yazılım geliştirme tekniği tanımlanamamaktadır [11]. Scrum anlaşılması kolay, karmaşık olmayan ama uzmanlaşması zor bir süreçtir. Bu süreçte öngörülebilirliği arttırmak ve riski kontrol etmek için yinelemeli ve artırımı bir yaklaşım kullanılır [9]. Artırımı bir yaklaşım kullanarak ürün aşama aşama inşa edilirken bir sonraki ürün parçası için öngörülebilirliğin artmasına bağlı olarak süreç içinde risklerin kontrol edilebilirliği de artmaktadır. Örneğin; yüz tanıma ile bir işlem yapacak mobil uygulama geliştirme aşamasında, ilk olarak yüz tanıma özelliği sistem üzerinde gerçekleştirilebilir. Daha sonraki aşamada detay işlemlerin nasıl gerçekleştirileceği artırımı bir şekilde sisteme eklenebilir.

Scrum süreci belirli bir düzen ile kendi içinde Koşu, Ürün Gereksinim Listesi, Koşu Gereksinim Listesi, Koşu İş Takip Grafiği, Koşu Hedefi ve Kullanıcı Hikayeleri gibi birçok kavram barındırmaktadır (Şekil 1).



Şekil 1. Scrum Süreci (Scrum Process) [8]

Scrum'da yer alan bu kavramlar, Scrum Sürecindeki Genel Kavramlar, Scrum Sürecindeki Roller, Scrum Sürecindeki Toplantılar olarak özetle aşağıdaki gibi tanımlanmaktadır:

3.1.1.Scrum Sürecindeki Genel Kavramlar (General Terms for the Scrum Process)

Koşu (Sprint): 1 haftadan 4 haftaya kadar süreleri olan yinelenmeli periyotlar “Koşu” olarak adlandırılmaktadır [1].

Ürün Gereksinim Listesi (Product Backlog): Temel olarak gereksinimlerin, Kullanıcı Hikâyelerinin ve özelliklerin bir listesidir [12] ve üründe yapılacak herhangi bir değişiklik için gereksinimlerin kaynağı olan bir liste şeklindedir [9].

Koşu Gereksinim Listesi (Sprint Backlog): Geliştirme Ekibinin “Koşu Hedefini” gerçekleştirmek için yapmaları gerekenleri tanımlayarak tüm işlerin görünür olmasını sağlayan bir liste şeklindedir [9].

Koşu İş Takip Grafiği (Sprint Burndown Chart): Bu grafik, ekibin her görevi tamamlaması için verdiği saat gün taahhüdünü gösterir ve ekibin ilerlemesi için bir ölçü olarak Scrum Uzmanı tarafından takip edilir [1].

Koşu Hedefi (Sprint Goal): Koşu Hedefi, Ürün Gereksinim Listesinin gerçekleştirilmesi için Koşu içinde uygulanacak olan ve Geliştirme Ekibine ürünü niçin geliştirdikleri konusunda rehberlik sağlayan bir amaçtır [9]. Koşu Hedefinin cevaplaması gereken temel soru “Niçin bu koşuyu gerçekleştiriyoruz” şeklindedir [12].

Kullanıcı Hikayeleri (User Stories): Gereksinimler genellikle Kullanıcı Hikayeleri oluşturularak açıklanmaktadır. Kullanıcı Hikayeleri müşterinin anlatımı ile oluşturulan, Ürün Sahibinin ne istediğini [1] ve sonuçta görmek istediğini ifade eden [12] ürün ile kullanıcının etkileşiminin basit bir açıklamasını içeren [13] bir liste şeklindedir. Kullanıcı Hikayeleri projede yapılacak görevleri ifade etmez, bu yüzden proje sürecindeki kişilerin ne yapacağını belirten görevlerden oluşmamaktadır [1].

3.1.2.Scrum Sürecindeki Roller (Roles in Scrum Process)

Scrum Ekibi; Ürün Sahibi, Scrum Uzmanı ve Geliştirme Ekibi olmak üzere 3 ana rolden oluşmaktadır [1]:

Ürün Sahibi (Product Owner): Geliştirilecek ürüne ihtiyacı olan ve işin gerekliliklerini başarmaktan sorumlu kişidir. Ürün Sahibi, Ürün Gereksinim Listesinde detaylandırılan ürünün vizyonunun [1], sorumluluğuna sahiptir [1]. Proje süreci ilerledikçe bu vizyon netleşir [14].

Geliştirme Ekibi (Development Team): İşlevselliğin geliştirilmesi sorumluluğuna sahip [14] ve yazılımı geliştiren 3 kişiden 15 kişiye kadar olan bir ekiptir [1].

Scrum Uzmanı (Scrum Master): Süreç boyunca takımı yönlendiren ve sorunları çözmek için ekibe yardımcı olan kişidir [1].

3.1.3.Scrum Sürecindeki Toplantılar (Meetings in the Scrum Process)

Koşu Planlama Toplantısı (Sprint Planning Meeting): Koşu boyunca gerçekleştirilecek olan işlerin planlandığı toplantıdır [9].

Günlük Scrum Toplantısı (Daily Scrum Meeting): Geliştirme Ekibinin gelecek 24 saat için bir plan oluşturması ve aktivitelere senkronize olmalarını sağlamak için 15 dakikadan oluşan olaylardır. Bu toplantılar ile bir önceki toplantıdan sonra yapılan çalışmalar görülür [9], yapılan iş miktarı [13] ve bir sonraki toplantıya kadar yapılacak çalışmalar tahmin edilir [9] ve bu şekilde ekip üyelerinin proje sürecindeki bilgiler hakkında fikir sahibi olmaları sağlanır [15].

Koşu Değerlendirme Toplantısı (Sprint Review Meeting): Gerçekleştirilen ürünü incelemek ve ihtiyaç duyulan Ürün Gereksinim Listesini düzenlemek için her Koşu sonunda gerçekleştirilen toplantılardır. Bu toplantı sırasında, Scrum Ekibi ve paydaşları Koşu sürecinde yapılan işler hakkında işbirliği içinde tartışmaktadır [9] ve paydaşlar Geliştirme Ekibinin ne yapıyor olduğunu öğrenmektedir [12].

Koşu Süreç Değerlendirme Toplantısı (Sprint Retrospective Meeting): Scrum Ekibinin kendisini değerlendirmesi için bir fırsat oluşturan ve sonraki Koşu sürecini iyileştirmek için plan oluşturulmasını sağlayan toplantıdır [9].

Bu çalışma kapsamında, Scrum metodu kullanılarak bir mobil uygulama geliştirme süreci gerçekleştirilmiştir.

4. MATERYAL VE METOD (MATERIALS AND METHODS)

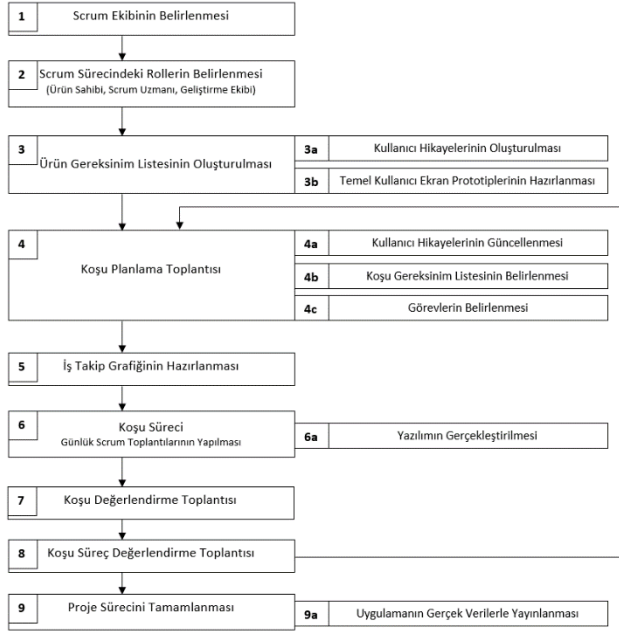
Scrum (2014) metodu ile geliştirilen mobil uygulama; verilerin sağlandığı veritabanı, veritabanı ile uygulama arasında iletişimi sağlayan servisler ve mobil uygulama olmak üzere üç kısımdan oluşmaktadır.

Veritabanı oluşturma kısmında, Microsoft SQL Server Management Studio platformu kullanılmıştır. Mobil uygulama geliştirme sürecinde Eclipse platformu üzerinde Java Tabanlı Android Yazılım Geliştirme dili kullanılmıştır. Bu iki aracın birbiri arasında iletişimi sağlayacak servislerin oluşturulması aşamasında ise Visual Studio platformu üzerinde C# ve Asp.Net yazılım geliştirme dili kullanılmıştır. Ayrıca süreç yönetimi için, Mock Up Builder [16] ve Ontime Scrum Planlama [17] yazılımlarının deneme sürümleri kullanılmıştır.

5. MOBİL UYGULAMA GELİŞTİRME SÜRECİNDE SCRUM'IN UYGULANMASI (IMPLEMENTATION OF SCRUM IN MOBILE APPLICATION DEVELOPMENT PROCESS)

Bu çalışma kapsamında; İstanbul Üniversitesi, İstanbul Tıp Fakültesi İlaç Araştırmaları Birimi için Android

tabanlı bir mobil uygulama Scrum metodu ile geliştirilmiştir. Çevik Metodolojiler içerisinde, geliştirilecek mobil uygulama için yazılım geliştirme sürecinde müşteri ihtiyaçlarını iyi analiz etmeyi sağlayan ve proje sürecinde herkesin aktif olmasına imkan veren Scrum metodu tercih edilerek, uygulama geliştirme sürecine başlanmıştır. Bu süreçte Scrum metodunun özellikleri kullanılarak mobil uygulama geliştirme süreci tamamlanmıştır. Bu çalışma kapsamında uygulanan Scrum metodu ile mobil uygulama geliştirme süreci Şekil 2'de özetlenmiştir.



Şekil 2. Scrum Metodu ile Mobil Uygulama Geliştirme Süreci (Mobile Application Development Process with Scrum Method)

Scrum metodu ile proje süreci başlarken ilk aşamada Scrum Ekibi oluşturulmuştur (1). Scrum Ekibinin oluşturulmasından sonra uygulama geliştirme sürecinde Ürün Sahibi, Scrum Uzmanı ve Geliştirme Ekibi'nin kimler olacağı belirlenmiştir (2). Bu aşamadan sonra Ürün Sahibi ile birlikte çalışarak, Ürün Sahibinin ihtiyaçları ve analizleri değerlendirilmiş ve Ürün Gereksinim Listesi belirlenerek (3), Kullanıcı Hikayeleri oluşturulmuş (3a) ve temel kullanıcı ekran prototipleri hazırlanmıştır (3b). Sürekli güncellenen Ürün Gereksinim Listesinden öncelikli olarak hangi ihtiyaçların karşılanacağını belirlemek için Koşu Planlama Toplantıları gerçekleştirilmiş (4), Ürün Sahibinin ihtiyaçlarını ifade eden Kullanıcı Hikayeleri güncellenmiş (4a), güncellenen Kullanıcı Hikayeleri seçilerek Koşu Gereksinim Listesi belirlenmiş (4b) ve Kullanıcı Hikayelerine göre görevler oluşturulmuştur (4c). Koşu Gereksinim Listesinin belirlenmesinden sonra Scrum Uzmanı tarafından, yapılan işlerin takibini kolaylaştırmak ve zaman içindeki performansı görmek amacıyla İş Takip Grafiği hazırlanmıştır (5). Bu aşamalardan sonra Koşu Süreci (6) başlatılmıştır. Ekibin kendi kendine organize olmasını sağlayan Günlük Scrum Toplantıları ile Koşu süreci

desteklenerek yazılım ve uygulama gerçekleştirilmiştir (6a). Bu süreç sonunda Koşu Değerlendirme Toplantısı (7) yapılarak uygulamanın o ana kadar yapılan kısmı sunulmuş ve geri bildirimler alınmıştır. Koşu Süreç Değerlendirme Toplantısında (8), süreç içinde Geliştirme Ekibini olumlu veya olumsuz etkileyen yönler değerlendirilmiştir. Koşu süreci yinelemeli olarak 3 kez gerçekleştirildikten sonra proje süreci tamamlanmıştır (9) ve ürün uygulama marketinde yayınlanmıştır (9a).

5.1. Ürün Gereksinim Listesinin Oluşturulması (Creating Product Backlog) (Şekil 2,3)

Ürün Sahibi önceden hazırladığı dokümantasyon üzerinde istediği ürünü genel olarak ifade ederken, Scrum Uzmanı ve Geliştirme Ekibi sorular sorarak Kullanıcı Hikayelerinin (Figure 2, 3a) oluşturulmasını sağlamıştır. Oluşturulan Kullanıcı Hikayelerinden örnekler şu şekildedir;

•Ben bir kullanıcı olarak, uygulamayı ilk açtığımda İstanbul Tıp Fakültesi yazan bir ekranla karşılaşmak istiyorum.

•Ben bir kullanıcı olarak, giriş ekranını geçtikten sonra, Konular, Öğretici Sorular ve Sık Sorulan Sorular şeklinde olan seçim yapabileceğim bir ekranın olmasını istiyorum.

5.2. Kullanıcı Ekran Prototiplerinin Hazırlanması (Creating User Interface Prototypes) (Şekil2, 3b)

Ürün sahibinin bu isteklerine göre Geliştirme Ekibi tarafından MockUp Builder aracı kullanılarak, Şekil 3'de görüldüğü gibi yazılımın tasarımsal prototipi hazırlanmıştır.

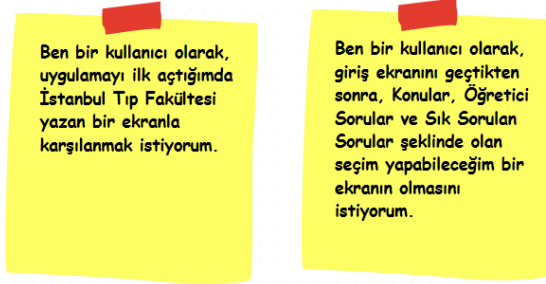


Şekil 3. Mobil uygulamanın tasarımsal prototip örnekleri (Examples of Conceptual Prototype to the Mobile Application)

5.3. Birinci Koşu Planlama Toplantısı (First Sprint Planning Meeting) (Şekil 2, 4)

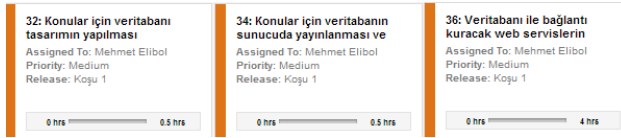
Bu toplantı öncesi Scrum Uzmanı ve Geliştirme Ekibi, Ürün Gereksinim Listesi üzerinde çalışarak Koşu Planlama Toplantısı öncesi hazırlık yapmışlardır.

Bu toplantı sırasında Ürün Gereksinim Listesinden Kullanıcı Hikayeleri güncellenmiş (Şekil 2, 4a) ve bu Kullanıcı Hikayelerinden (Şekil 4) hangilerinin Koşu sürecinde gerçekleştirileceğine karar verilerek Koşu Gereksinim Listesine (Şekil 2, 4b) aktarılmıştır.



Şekil 4. Birinci Koşu için Kullanıcı Hikayeleri Örnekleri (Examples of User Stories for the First Sprint)

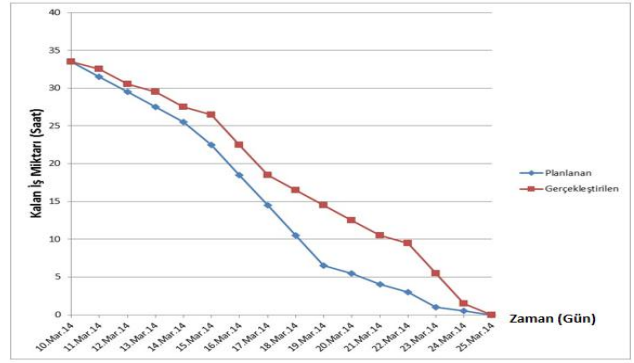
Kullanıcı Hikayelerini gerçekleştirilmek için yapılması gereken görevler (Şekil 2, 4c) ve bu görevlerin ne kadar birimde¹ gerçekleştirileceği Şekil 5'de yer alan örnekte görüldüğü gibi planlanmıştır.



Şekil 5. Birinci Koşu için Gerçekleştirilmesi Gereken Görevler (Tasks to be performed for the First Sprint)

5.4. Scrum Uzmanı Tarafından İş Bitim Grafiğinin Hazırlanması (Preparation of Sprint Burndown Chart by Scrum Master) (Şekil 2, 5)

Scrum Uzmanı tarafından Koşu sırasında işlerin ne kadarının yapıldığının rahat bir şekilde takip edilebilmesi için Koşu başlangıcında İş Takip Grafiği (Şekil 6) oluşturulmuş ve Geliştirme Ekibi ile paylaşılmıştır. Bu grafik üzerinde saat bazında yapılan işler çizilerek zamanlama hedefi ile ne kadar doğru orantılı olarak gidildiğinin görülmesi sağlanmıştır.



Şekil 6. Birinci Koşu Sırasında Takip Edilen İş Takip Grafiği (Sprint Burndown Chart for the First Sprint)

5.5. Birinci Koşu Süreci (First Sprint Process) (Şekil 2, 6)

Birinci Koşu sürecinde; veri tabanı tasarımı gerçekleştirilmiş ve sunucuya aktarılmış, web servisler oluşturulmuş ve sunucuya aktarılmış, canlı ortamda çalışan servisler aracılığıyla, tekrar kullanılabilir yapılar kullanılarak, uygulama geliştirme sürecine başlanmıştır (Şekil 2, 6a).

Birinci Koşu sürecinde ortaya çıkan bazı önemli noktalardan bir tanesi şu şekildedir;

- Ürün Sahibinden gelen dönütler sonucunda yönlendirme ekranında İletişim ve Uygulama Hakkında seçeneklerinin olmasına karar verilmiştir.

5.6. Birinci Koşu Sürecinde Günlük Scrum Toplantıları (Daily Scrum Meetings during First Sprint Process) (Şekil 2, 6)

Günlük Scrum Toplantıları gerek yüz yüze olsun, gerek iletişim teknolojilerinin getirdiği yeniliklerden yararlanılarak çeşitli yöntemlerle gerçekleştirilmiştir.

5.7. Birinci Koşu Değerlendirme Toplantısı (First Sprint Review Meeting) (Şekil 2, 7)

Geliştirme Ekibi tarafından uygulamanın o ana kadar tamamlanmış kısmının sunumu yapılmış ve Ürün Sahibinden dönütler alınmıştır. Örneğin;

- Uygulamanın ilk açıldığı ekranda, logo, belirli bir süre (3-5 saniye kadar) ekranda kaldıktan sonra Devam butonuna gerek kalmadan da yönlendirme ekranına geçilebilmelidir.

¹ Koşu süresince iş hacmi birim kavramı ile ifade edilmektedir. Uygulama geliştirme süresince birim kavramına karşılık olarak iş, saat olarak ifade edilmiştir.

5.8. Birinci Koşu Süreci Değerlendirme Toplantısı (First Sprint Retrospective Meeting) (Şekil 2, 8)

Bu toplantı sırasında üzerinde durulan noktalara ait örneklerden bir tanesi aşağıdaki şekildedir;

- Zaman Yönetimi konusunda ekibin daha dikkatli olması gerektiği üzerinde durulmuştur.

5.9. İkinci ve Üçüncü Koşu (Second and Third Sprints)

İkinci ve Üçüncü Koşu süreçleri kısaca aşağıdaki şekilde gerçekleşmiştir. Detayları bir önceki bölümde anlatılan birinci koşu sürecine benzerlik göstermektedir.

İkinci Koşu süreci 8 günlük sürede, her gün Günlük Scrum Toplantıları gerçekleştirilerek, Kullanıcı Hikayeleri temel alınarak hazırlanmış Görev Kartları ve Scrum Uzmanı tarafından hazırlanan İş Takip Grafiği takip edilerek gerçekleştirilmiştir. İkinci Koşu süreci Koşu Değerlendirme Toplantısı ve Koşu Sürec Değerlendirme Toplantısı gerçekleştirilerek tamamlanmıştır.

Üçüncü Koşu süreci 10 günlük sürede, İş Takip Grafiği takip edilerek ve Görev Kartlarındaki görevler gerçekleştirilerek tamamlanmıştır. Üçüncü Koşu sonunda Geliştirme Ekibi uygulamanın son halinin sunumunu yapmıştır ve artık uygulamanın gerçek verilerle Google Markette yayınlanmasına karar verilmiştir.

5.10. Ürünün Yayınlanması ve Proje Sürecinin Kapatılması (Launching the Production to the Market and Closing the Project Process) (Şekil 2, 9)

Ürün Sahibinden gelen gerçek veriler sisteme girildikten sonra, Ürün Sahibinin ve kullanıcıların gerçek verilerle sistemi test etmeleri sağlanmıştır. Gerçek verilerle kullanım sonucunda oluşan geri bildirimler doğrultusunda uygulama üzerinde güncellemeler yapılmıştır. Ürün Sahibinden onay alınarak uygulama Google Markette yayınlanmıştır (Şekil 2, 9a).

6. TARTIŞMA VE SONUÇ (DISCUSSION AND RESULT)

Bu çalışma sürecinde değişime en iyi şekilde cevap verebilmek amacıyla geliştirilen mobil uygulama çevik metodolojilerden olan Scrum metodu takip edilerek gerçekleştirilmiştir. Bu şekilde uygulama geliştirme sürecinde değişikliklere karşı duyarlılık oluşmuş ve değişikliklere en kısa sürede cevap verebilmek mümkün olmuştur.

Scrum Metodu proje sürecini işin ne kadarının bittiğinin görülmesi için İş Takip Grafiğinin oluşturulmasını zorunlu kılmaktadır[4]. Buna uygun olarak mobil uygulama geliştirme sürecinde İş Bitim Grafikleri oluşturulmuş, Geliştirme Ekibinin süreç içinde, her gün

işin ne kadarının yapıldığını görmeleri sağlanarak sürecin şeffaf olması sağlanmıştır. Bunun proje sürecinin takip edilmesinde son derece önemli olduğu, Scrum ile yürütülen projelerde mutlaka yapılması gerektiği sonucuna varılmıştır.

Schwaber ve Sutherland [9] Scrum Metodunda etkileşim ve iletişim kavramlarının önemi üzerinde durmaktadır. Buna uygun olarak, mobil uygulama geliştirme sürecinde Ürün Sahibi sürekli ulaşılabilir durumda olmuş, oluşan olumlu etkileşim ve pozitif iletişim ile değişikliklere karşı duyarlılık oluşmuş ve soru işareti olan noktaların hızlı şekilde çözülmesi sağlanmıştır. Özellikle mobil uygulama sektörü web sektörüne göre yeni sayılabilecek bir alan olduğu için önceden öngörülememiş ve bilgisayar teknolojilerine göre düşünülmüş noktaların (Ör: mobil işletim sistemi versiyonlarında kodların çalışmaması) hızlı şekilde çözülerek, mobil uygulama geliştirme sürecinin sağlıklı devam etmesi sağlanmıştır.

Yazılım geliştirme sürecinde önemli noktalardan birisi yazılımın süreç içinde test edilmesini sağlamaktır. Bu çalışma sürecinde Beck ve arkadaşlarının [4] belirttiği gibi mobil uygulama çalışır durumda tutulduğu için, uygulamanın sürekli test edilmiş, bu şekilde hataların anında tespit edilmesi ve çözümün hızlı şekilde bulunması sağlanmıştır. Bilgisayar teknolojilerinden farklı olarak birçok mobil cihaz ve mobil işletim sistemi versiyonunun olması ve farklı mobil cihaz ve işletim sistemi üzerinde farklı hataların çıkabileceği ve o cihaza özel çözüm üretilmesi gerektirebileceği için özellikle mobil uygulama geliştirme sürecinde daha önemli olduğu düşünülmektedir.

Scrum süreci denetim ve adaptasyonu başarı ile gerçekleştirmek için farklı toplantıların yapılmasını zorunlu kılmaktadır [9]. Bu çalışma kapsamındaki uygulama geliştirme sürecinde bu toplantılar yüz yüze ve/veya uzaktan haberleşme yöntemleriyle gerçekleştirilmiştir. Bunun sonucunda, uzaktan haberleşme yöntemleri ile yapılan toplantıların yerine, yüz yüze iletişimin sağlandığı toplantıların daha verimli geçtiği görülmüştür. Ayrıca Ürün Sahibi ile yapılan yüz yüze toplantılar sonucunda, Ürün Sahibinin ifade etmek istediklerinin kağıt kalem kullanarak, basit şekilde çizilmesinin, ne yapılacağı belirlenmesine ve prototipler hazırlanmasına büyük kolaylık sağladığı görülmüştür. Uzaktan haberleşme yöntemleri ile yapılan toplantılarda aynı verim sağlanamamıştır.

Ürün Gereksinim Listesinin belirlendiği ilk toplantı öncesi Scrum Uzmanı ve Geliştirme Ekibinin o alanda yapılmış çalışmaları inceleyerek, toplantıya gelmesinin bir gereklilik olduğu görülmüştür. Özellikle Scrum Uzmanı ve Geliştirme Ekibinin yabancı olduğu alanlarda bir iş geldiğinde bunun daha da önemli olabileceği düşünülmektedir. Klinik Araştırmalar için mobil uygulama geliştirme proje süreci başlamadan önce Scrum Uzmanı ve Geliştirme Ekibi bu alanda yapılan çalışmaları

incelemiştir. Bu şekilde, Scrum sürecinde yer alan tüm rollerin ortak bir dili konuşması sağlanmıştır.

Mobil uygulama geliştirme sürecinde, müşteriler mobil uygulama dünyasındaki bazı kısıtları düşünmeden ne istediklerini anlatmakta ve web tabanlı sistemlerdeki gibi özellikleri sistemlere ekletmeye çalışabilmektedirler. Bunun için Ürün Sahibinin de mobil uygulamaların web teknolojilerinden farklı olduğunu ve isteklerini ona göre iletmesi için örnek mobil uygulamaları incelemesinin yararlı olacağı görülmüştür. Bu çalışma sürecinde bilgilerin eklenmesi sırasında, ürün sahibi bir sayfaya çok uzun bilgilerin eklenmesini istemiş. Scrum Uzmanı devreye girerek, mobil uygulamalarda tasarımsal olarak daha iyi olacak ve kullanılabilirliği artıracığı için bir sayfada daha az bilgi verilmesinin uygun olacağını belirtmiş ve bilgilerin mobil uygulama üzerinde parça parça yapılandırılmasını sağlamıştır.

Scrum süreci yürütülürken bu sürecin çeşitli yazılımlarla desteklenmesinin önemli olduğu düşünülmekte ve süreci takip etmek ve ekran tasarımlarını yapmak için bir yazılımın kullanılması önerilmektedir. Çalışmada kullanılan web tabanlı OnTimeScrum Planlama[17] yazılımı sürecin ihtiyaçlarını karşılayarak sürecin verimli geçmesini sağlamıştır. Yine süreçte kullanılan MockUp Builder[16] ekran prototip hazırlama aracı, ekran tasarımlarının hazırlanmasında kolaylık sağlamıştır. Süreç içinde bu yazılımların deneme sürümleri kullanılmıştır.

Jeong ve ark [7] belirttiği gibi tüketicinin görüşünü yansıtan ve pazar ihtiyacını karşılayan mobil uygulamaların kar sağlama için mümkün olduğu kadar hızlı şekilde pazara çıkması önemli olmaktadır. Bu çalışma sürecinde de buna uygun olarak Scrum Metodu ile gerçekleştirilen mobil uygulama yaklaşık 2 ay gibi bir sürede Google Play Markette kullanıma sunulmuştur.

Sonuç olarak, bu çalışma kapsamında çevik metodolojilerden birisi olan Scrum kullanılarak mobil uygulama geliştirilmiştir. Değişim sürecine hızlı adaptasyon, müşteri isteklerine hızlı cevap, kısa sürede markete çıkma, test süreçlerini daha hızlı bir şekilde gerçekleştirme, uygulamanın sürekli çalışır olması, bireyler arasındaki etkileşim, Ürün Sahibi ile işbirliği sunulması gibi artılarından dolayı mobil uygulama geliştirme sürecinde Scrum Metodunun tercih edilmesi önerilmektedir. Ancak projeye başlamadan önce Scrum sürecinin ve çevik manifesto felsefesinin tüm ekip tarafından anlaşılması hatta benimsenmiş olmasının ve mobil uygulama geliştirme sürecinde çeşitli sınırlılıklar ve yeniliklerin olduğunun ekip tarafından kavranmış olmasının önemli olduğu düşünülmektedir.

Bu çalışmadan elde edilen çıktıların metodoloji seçiminde kararsız kalan araştırmacılar ve Scrum ile mobil uygulama geliştirecek kişiler için önemli bir kaynak olduğu düşünülmektedir. Özellikle Scrum Metodunun mobil oyun sektöründe aşamalı oyunlarda uygulanmasının, pazar rekabetinde avantaj sağlayacağı

düşünülmektedir. Her aşama bir Koşu sürecinde bitirilerek, bu Google Markette yayınlanabilir. Bu şekilde rekabetin olduğu pazarda uygulamanın mümkün olan en kısa sürede pazara çıkması sağlandığı için kâr etme oranı artırılabilir. Ayrıca Scrum sürecinde her Koşu sonucunda oluşan ürün parçasının yayınlanması ile uygulamanın Google Markette arama sonucunda ön sıralarda yer alması sağlanabilir. Bu da tam rekabet piyasasında ürün açısından önemli bir avantaj olarak görülmektedir.

KAYNAKLAR (REFERENCES)

- [1] C. Scharff, R. Verma, "Scrum to support mobile application development projects in a just-in-time learning context", **Proceedings of the 2010 ICSE Workshop on Cooperative and Human Aspects of Software Engineering**, Cape Town, New York, Association for Computing Machinery, ISBN:978-1-60558-966-4, 25-31, 01-08 Mayıs 2010.
- [2] H.K. Flora, S.V. Chande, "A Review and Analysis on Mobile Application Development Processes Using Agile Methodologies", *International Journal of Research in Computer Science*, 3 (4), 8-18, 2013.
- [3] V. Rahimian, R. Ramsin, "Designing an Agile Methodology for Mobile Software Development: A Hybrid Method Engineering Approach", **2nd International Conference on Research Challenges in Informations Science**, Marrakech, Morocco, ISBN: 978-1-4244-2273-9, 337-342, 3-6 Haziran 2008.
- [4] Internet: K. Beck, M. Beedle, A. Bennekum, A. Cockburn, W. Cunningham, M. Fowler, J. Grenning, J. Highsmith, A. Hunt, R. Jeffries, J. Kern, B. Marick, R. C. Martin, S. Mellor, K. Schwaber, J. Sutherland, D. Thomas, Manifesto for Agile Software Development, <http://agilemanifesto.org/>, 25.02.2014.
- [5] M. Elibol, and Ç. S. Erol, "Mobile Application Development With Agile Methodology", **Economics and Business Communication Challenges**, Editör: Dyduch, W. And Pankowska, M. , House of the University of Economics in Katowice, Katowice, Poland, 72-86, 2014.
- [6] P. Abrahamsson, A. Hanhineva, H. Hulkko, T. Ihme, J. Jaalinoja, M. Korkala, J. Koskela, P. Kyllonen, P. Salo, "Mobile-D:an agile approach for mobile application development", **Proceeding OOPSLA '04 Companion to the 19th annual ACM SIGPLAN conference on Object-oriented programming systems, languages, and applications**, Vancouver, BC, Canada, ISBN:1-58113-833-4, 174-175, 24-28 Ekim 2004.
- [7] Y.J. Jeong, J.H. Lee, G.S. Shin, "Development Process of Mobile Application SW Based on Agile Methodology", **10th International Conference on Advanced Communications Technology**, Gangwon-Do, Institute of Electrical and Electronics Engineers (IEEE), ISBN: 978-89-5519-136-3, 362-366, , 17-20 Şubat 2008.
- [8] J. Highsmith, **Agile Software Development Ecosystems**, Addison Wesley, Boston, ISBN: 0-201-76043-6, 2002.
- [9] K. Schwaber, J. Sutherland, **The Definitive Guide to Scrum: The Rules of the Game**, Temmuz 2013.
- [10] K. Vlaanderen, S. Jansen, S. Brinkkemper, E. Jaspers, "The agile requirements refinery:Applying SCRUM principles to software product management", *Information and Software Technology*, 53 (1), 58-70, 2011.

[11] P. Abrahamsson, O. Salo, J. Ronkainen, J. Warsta, **Agile Software Development Methods: Review and Analysis**, VTT Publications 478 Espoo, Finland, 2002.

[12] H. Kniberg, **Scrum and XP from the Trenches**, C4 Media Inc, United States of America , ISBN:978-1-4303-2264-1, 2007.

[13] Internet: P. Nicolas, Introduction to SCRUM Agile Process for Global Software Development, http://www.pnexpert.com/files/SCRUM_for_Global_Development.pdf, 25 Nisan 2014.

[14] K. Schwaber, **Agile Project Management with Scrum**, Microsoft Press, Washington, , ISBN:0-7356-1993-X, 2004.

[15] M. Beedle, M. Devos, Y. Sharon, K. Schwaber, J. Sutherland, "SCRUM: An extension pattern language for hyperproductive software development", **Pattern Languages of Program**, 4, 637-651, 1999.

[16] Internet: <http://mockupbuilder.com>, 27.03.2017

[17] Internet: <https://www.axosoft.com>, 27.03.2017

Büyük Veride Kişi Mahremiyetinin Korunması

Can EYÜPOĞLU^{1*}, Muhammed Ali AYDIN², Ahmet SERTBAŞ², Abdül Halim ZAIM³, Onur ÖNEŞ⁴

¹Bilgisayar Mühendisliği Bölümü, İstanbul Ticaret Üniversitesi, İstanbul, Türkiye

²Bilgisayar Mühendisliği Bölümü, İstanbul Üniversitesi, İstanbul, Türkiye

³Bilişim Teknolojileri Uygulama ve Araştırma Merkezi, İstanbul Ticaret Üniversitesi, İstanbul, Türkiye

⁴Bilişim ve İletişim Teknolojileri, Turkcell Teknoloji, İstanbul, Türkiye

ceyupoglu@ticaret.edu.tr, aydinali@istanbul.edu.tr, asertbas@istanbul.edu.tr, azaim@ticaret.edu.tr, onur.ones@turkcell.com.tr

(Geliş/Received:26.09.2016; Kabul/Accepted:03.02.2017)

DOI: 10.17671/gazibtd.309301

Özet—Büyük verinin ortaya çıkışı bilgi güvenliği ve klasik güvenlik tedbirleri için kullanılan koruma modelleri için yeni zorluklara neden olmaktadır. Bu çalışmada büyük veri güvenliği ve büyük veride kişi mahremiyetinin korunmasına yönelik literatürde var olan çalışmalar özetlenmiştir. Buna ek olarak büyük veri kaynaklarının neler olduğu, büyük veri sistemlerini korumak için gerekli olan araçlar ve bu sistemlerin güvenliğinin sağlanmasında karşılaşılan zorluklar açıklanmıştır.

Anahtar Kelimeler—Büyük veri, büyük veri güvenliği, kişi mahremiyeti

Preserving Individual Privacy in Big Data

Abstract—The emergence of big data causes new difficulties for protection models used for information security and conventional security precautions. In this study, existing studies in the literature related to big data security and preserving individual privacy in big data are summarized. Besides, big data sources, necessary means to protect big data systems and the difficulties in ensuring the security of these systems are explained.

Keywords—Big data, big data security, individual privacy

1. GİRİŞ (INTRODUCTION)

Büyük veri; sosyal medya paylaşımları, GSM operatörlerinden elde edilen arama kayıtları, fotoğraf, video, blog ve log dosyaları gibi farklı kaynaklardan elde edilen verilerin anlamlı ve işlenebilir hale dönüştürülmüş biçimidir. Şu anda yapısal veriler ilişkisel veritabanlarında tutulmaktadır ve birçok şirket günlük olarak terabaytlarca veriyi yönetmek zorundadır. Geleneksel altyapı bu tür veriler üzerinde işlemler gerçekleştirmek ve bunların analitik verilerini oluşturmak için eksik kalmaktadır. İlişkisel veritabanlarını kullanan şirketler veri madenciliği yöntemleri ile elde ettikleri verileri kullanarak karar almaktadır. Fakat günümüzde bu tip bir çözüm yeterli değildir. Firmalar yapısal olmayan büyük veriyi oluşturmak, analiz etmek ve saklamak zorundadır. Klasik veritabanları sürekli artan bu boyuttaki veriyi tutmakta başarılı olamamaktadır. Ayrıca büyük verinin ortaya çıkışı kişi mahremiyetinin korunmasına yönelik yeni zorluklara neden olmaktadır. Geleneksel veritabanları için birçok başarılı kimliksizleştirme ve veri güvenliği çözümü bulunmaktadır. Ancak büyük veri üzerinde geliştirilmeye başlanan bu tip çözümler daha emekleme aşamasındadır.

Organizasyonlar güvenliği ve mahremiyeti sağlamak için çeşitli kimliksizleştirme (de-identification) yöntemleri kullanmaktadır. Güvenliği ve mahremiyeti garanti altına almak için en yaygın çözüm sözlü ve yazılı taahhütlerdir. Fakat bu çözümün başarısız olduğu görülmüştür. Şifreler, denetimli erişim ve iki-faktörlü kimlik doğrulama; veri dinamik ve dağıtık veri sistemlerinde paylaşıldığı ve bir araya getirildiğinde güvenliği ve gizliliği sağlamak için düzenli olarak kullanılan teknik çözümlerdir ancak düşük seviyeli. Daha gelişmiş teknik çözüm ise kriptografidir. AES ve RSA tanınmış şifreleme algoritmalarıdır. Son günlerdeki açıklamalar NSA (National Security Administration)'nın mevcut Internet şifreleme algoritmalarını kırmanın yollarını bulduğunu göstermektedir [1, 2].

Güvenlik duvarları (firewalls), güvenli soket katmanı, taşıma katmanı güvenliği gibi sanal engeller veriye erişimi sınırlamak için tasarlanmıştır. Bu teknolojilerin hepsi kırılabilir. Bu sebeple sürekli olarak gözlemlenmelidir ve gerektiğinde onarılmalıdır. İzleme, gözlemlenme veya denetleme yazılımları güvenliğe uygunluğu garanti altına almak için bireysel kullanıcıların veri akış ve ağ erişim geçmişlerini temin etmek için tasarlanmıştır. Bu teknolojinin kısıtlaması büyük ölçekli

veya dağıtık veri sistemlerinin gerçekleştirilmesinin zor ve maliyetli olmasıdır. Çünkü bulguların okunması ve yorumlanması gerekmektedir. Ayrıca yazılım veriyi korumadan ziyade bireysel davranışları gözlemlemede kullanılabilir. Sonuç olarak geleneksel kimliksizleştirme teknikleri yaygın kullanımları nedeniyle büyük veri çağında uygulanabilir değildir. Büyük veri güvenliği ve gizliliğini sağlama görevleri bilgi arttıkça daha zor olmaktadır. Bilgisayar uzmanları isimsizleştirilmiş (anonymized) verinin sıklıkla tekrardan tespit edildiğini ve belirli bir bireye dayandırıldığını sürekli olarak göstermektedir [1, 3].

Sosyal ağ büyük veride önemli bir veri kaynağıdır. Fakat bu veri büyük oranda kullanıcıların özel verisini içermektedir. Sosyal ağ anonimlik koruması kullanıcı ID'si için anonimliği gerçekleştirmektedir. Sosyal ağ anonimlik korumasının en önemli sorunu saldırganın anonim bir kullanıcı olabmesidir. Bu anonim kullanıcı, kullanıcıların halka açık olan bilgileri ve kullanıcılar arasındaki bağlantıları kullanarak sonuç çıkarabilir [1]. Örnek olarak çeşitli bağlantı tahmin algoritmaları [4-8]'de öne sürülmüştür.

Çalışmanın ikinci bölümünde büyük veri sistemlerini korumak için gerekli olan araçlardan bahsedilecektir. Üçüncü bölümde büyük veri sistemlerinin güvenliğinin sağlanmasındaki zorluklara değinilecektir. Dördüncü bölümde büyük veri kaynaklarına yer verilecektir. Beşinci bölümde ise büyük veri güvenliği için geliştirilen yöntemlerden bahsedilecek ve altıncı bölümde sonuç verilerek makale sonlandırılacaktır.

2. BÜYÜK VERİ SİSTEMLERİNİ KORUMAK İÇİN GEREKLİ OLAN ARAÇLAR (NECESSARY MEANS FOR PROTECTION OF BIG DATA SYSTEMS)

Büyük veri bilgi güvenliği ve geleneksel güvenlik tedbirleri için benimsenen koruma ideolojisi için yeni tehditlere yol açmaktadır. Büyük veri güvenliği konularını çalışan bir grup olan CSA (Cloud Security Alliance, Bulut Güvenlik Antlaşması) son olarak büyük veri sistemlerini korumak için gerekli olan araçları listeleyen bir doküman hazırlamıştır [9, 10]:

1. Dağıtık (distributed) programlama çerçeveleri (framework) için güvenilir hesaplamalar,
2. İlişkisel olmayan veri depolama alanları için güvenlik uygulamaları,
3. Güvenilir veri depolama ve işlem logları,
4. Uç nokta giriş onaylama/filtreleme,
5. Gerçek zamanlı güvenlik görüntüleme,
6. Ölçeklenebilir ve birleştirilebilir gizlilik korumalı veri madenciliği ve matematiksel analiz,
7. Kriptografik zorunlu veri merkezli güvenlik,
8. Granüler erişim kontrolü,
9. Granüler hesap denetimleri,
10. Veri kaynağı (provenance).

CSA bu araçları dört gruba bölmüştür: altyapı güvenliği, veri koruma, veri yönetimi ve reaktif güvenlik.

Hadoop dağıtık programlama çerçevesine örnek olarak verilebilir. Hadoop Java ile yazılmış açık kaynak kodlu bir çerçevedir ve sıradan sunucular üzerinde büyük veri işlemeye yarar. HDFS (Hadoop Distributed File System, Hadoop Dağıtık Dosya Sistemi) birden çok bilgisayar üzerinde dağıtık olarak çalışan bir dosya sistemidir ve makineler üzerindeki dosya sistemlerini birbirine bağlayarak tek bir dosya sistemi gibi gözükmesini sağlar. İlişkisel olmayan veri depolama alanları yani büyük veri ortamları için geliştirilecek güvenlik uygulamaları, bu uygulamalarda yer alacak olan güvenilir veri depolama ve güvenilir işlem logları büyük öneme sahiptir. Ayrıca bu sistemlerde uç nokta girişlerin onaylanması, filtrelenmesi ve gerçek zamanlı olarak sistem görüntülenmesi gerekmektedir. Veri madenciliği uygulamalarının ve matematiksel analizlerin ise kişi mahremiyeti göz önünde bulundurularak yapılması gerekmektedir. Bunlara ek olarak kriptografik uygulamaların yer aldığı merkezi bir güvenlik birimine ihtiyaç duyulmaktadır. Granüler hesap erişimlerinin denetlenmesi, kontrolü ve veri kaynağının güvenilir olması büyük veri sistemlerini korumak için gerekli olan diğer araçlardır.

3. BÜYÜK VERİ SİSTEMLERİNİN GÜVENLİĞİNİN SAĞLANMASINDAKİ ZORLUKLAR (CHALLENGES IN PROVIDING SECURITY OF BIG DATA SYSTEMS)

Siber güvenlik problemleri için olan büyük veri analiz uygulamaları önemli derecede problemleri çözse de gerçek potansiyeli gerçekleştirmek için birkaç zorluğa değinmek gerekmektedir. Bu zorluklar şunlardır [9]:

1. Gizlilik (mahremiyet),
2. Büyük veri analizi APT (Advanced Persistent Threats, Gelişmiş Kalıcı Tehditler) algılama,
3. Yüksek performanslı kriptografi,
4. Güvenlik araştırması için büyük veri setleri,
5. Veri kaynağı problemi,
6. Güvenlik görselleştirme,
7. Nitelikli personel.

Büyük verinin ortaya çıkışı ile bilgi güvenliği ve geleneksel güvenlik tedbirleri için olan çözümler yetersiz kalmaktadır. Bu noktada kişilerin mahremiyeti büyük öneme sahiptir. Büyük veri sistemlerinde ortaya çıkabilecek olan kalıcı tehditlerin algılanması gerekmektedir ve gerekli noktalarda yüksek performanslı kriptografik uygulamalar kullanılmalıdır. Büyük veride veri kaynağının ve araştırmalarda kullanılacak olan veri setlerinin güvenilir olması çok önemlidir. Ayrıca büyük veri sistemlerini tasarlayan, geliştiren ve güvenlik problemlerini gözlemleyen kişilerin bilgi birikiminin yeterli düzeyde olması gerekmektedir.

Büyük veri için olan tehditler ve büyük veri altyapısı için düşünülmesi gerekenler şöyledir [11, 12]:

1. İşlemeyen verinin korunması,
2. Yönetimsel veri erişimi,
3. Yapılandırma ve yama (patch) yönetimi,
4. Uygulamaların ve düğümlerin kimlik doğrulamaları,
5. Denetim ve loglama,
6. İzleme, filtreleme ve engelleme,
7. API (Application Programming Interface, Uygulama Programlama Arayüzü) güvenliği.

Geliştirilecek olan büyük veri sistemlerinin altyapısında işlenmeyen yani sadece depolanan verinin korunması, verilere erişimin yönetilmesi ve var olan sistemin geliştirilmesi için eklenen yamaların yönetimi çok önemlidir. Bu sistemlerde uygulamaların ve düğümlerin kimlik doğrulamalarının yapılması, denetlenmesi ve log dosyalarının tutulması gerekmektedir. Ayrıca bu sistemler izleme, filtreleme, engelleme ve API güvenliği gibi mekanizmalara sahip olmalıdır.

4. BÜYÜK VERİ KAYNAKLARI (BIG DATA SOURCES)

Güvenlik analizinizi daha güvenli olması için veri kaynakları göz önünde bulundurulmalıdır. Güvenlik analizleri (analytics) için olan büyük veri kaynakları şunlardır [11, 13]:

1. Sistem-tabanlı veri: IP konumları (locations), klavye yazma veya mouse tıklama akış desenleri vb.,
2. Mobil-tabanlı veri: GPS konumları, ağ konumları vb.,
3. Ağın fiziksel fazlalığının zamanı ve konumu,
4. Hareket (ulaşım) verisi: hareket desenleri, kaynakları, hedefleri vb.,
5. Dış yetkisiz kaynaklardan gelen veri,
6. Kimlik verisi: kullanıcı ismi ve şifresi vb.,
7. OTP (One Time Passwords): online erişim için kullanılan tek seferlik şifre,
8. Kimlik doğrulamada (authentication) kullanılan kullanılan dijital sertifikalar,
9. Biyometrik kimlik saptama (identification) verisi: parmak izi, iris, ses tanıma (speech recognition),
10. Sosyal medya verisi: Facebook, Google Drive, Twitter vb.

Veriyi daha güvenli ve verimli hala getirmek için dikkat edilmesi gereken kaynaklar şunlardır [11]:

1. Ağ trafiği,
2. Ağ kaynakları,
3. Kullanıcı kimlik bilgileri,
4. Ağ sunucuları.

5. BÜYÜK VERİ GÜVENLİĞİ İÇİN GELİŞTİRİLEN YÖNTEMLER (METHODS DEVELOPED FOR BIG DATA SECURITY)

Büyük verinin gizliliği en zorlayıcı faktörlerden biri haline gelmiştir. Veri setlerinin dramatik bir şekilde artmasıyla depolama ve analiz ihtiyacı da artmaktadır. Bu sebeple güvenlik de artırılmalıdır. Çünkü yetkisiz bir kişi belirli bir kişinin verisini indirebilir ve ona zarar verebilir. Saldırganları takip etmek için farklı gözetleme (surveillance) araçları ve takip cihazları kullanılmaktadır. Fakat veri arttıkça güvenlik de artırılmalıdır. Genel olarak kişisel verinin toplanması, depolanması, analizi ve kullanımı güvenliğin her seviyesinde günlük hayatımızın bir parçası olmuştur. Bu problem için bir çözüm de desen gizlemedir (pattern hiding). Bu çözümde veri yalnızca analistin çözebileceği bir desen içerisinde saklanır. Bu veri setleri ayrıca veri toplayıcısı tarafından da tanınmaz olmalıdır. Aksi takdirde veri yetkisiz taraflara sızdırılabilir. Güvenli dağıtım madencilik (secure distributive mining) ayrıca çok sayıda veri seti üzerinde de yapılabilir. Burada madencilikte kullanılacak veri farklı gruplar arasında yatay veya dikey olarak parçalara ayrılır ya da dağıtılır. Bölümlenmiş veri paylaşılamaz ve özel kalmalıdır. Ancak birleştirilen veri üzerindeki madenciliğin sonuçları katılımcılar (participants) arasında paylaşılır [11].

Verinin k-anonimliği veya kimliksizleştirilmesi kullanıcıların mahremiyetini tehlikeye atmadan web üzerinden verilerin paylaşılması yeteneğini ifade etmektedir. Diğer bir deyişle k-anonimlik bilginin açığa çıkma riskini sınırlarken veri kullanımını maksimuma çıkaran bir güvenlik kavramıdır. Buradaki bilgi açık bir kimlik olmasa bile ilgili varlığın kimliğinin tespitine yol açmaktadır [14].

Büyük veri kullanımının artan başarısı ile birçok zorluk ortaya çıkmıştır. Araştırmacıların çözmeye çalıştığı ana problemler zamansızlık (timeless), ölçeklenebilirlik (scalability) ve gizlilik (privacy). Gizlilik koruma şu anda oldukça aktif bir araştırma alanıdır. Bu konu ile ilgili olarak birçok çalışma ve konsept vardır. Bu konseptlerden biri de kimliksizleştirme tekniğidir. Kimliksizleştirme hassas bilgiyi bulma ve silmeden oluşan özel bir alandır. Kimliksizleştirme kriptografi ve veri madenciliği gibi teknikleri kullanarak bilginin yerinin değiştirilebilir, şifreleyebilir ya da bilgiye bir gürültü ekleyebilir. Rahmani ve diğerleri tarafından yapılan çalışmada [14] CLONALG olarak bilinen özel bir bağışıklık sistemi kullanarak metin verilerini kimliksizleştiren yeni bir model ortaya konulmuştur. Bu yaklaşımda kimliksizleştirme (de-identification) ve tekrardan kimlik saptama (re-identification) olmak üzere iki ana aşama vardır. Kimliksizleştirme kullanıcıların verilerinin içerisinde yer alan kimliklerinin korunmasını sağlayan ana aşamadır. Bu aşama belirleyici (identifier) olarak düşünülen herhangi bir bilginin bulunması ve yerinin değiştirilmesi için kullanılmaktadır. Bu aşama da beş adımdan oluşmaktadır: tokenların ayırma (tokenization), kodlama (codification), tespit etme (detection), depolama

(storage) ve yer değiştirme (replacement). Tokenlarına ayırma adımı bir kelime dizisi içerisinde kimliksizleştirme yapmak için kullanıcıların bilgisini ve veriyi ayrıştırırken bir kelime paketi kullanılmaktadır. Kodlama adımı kimliklerdeki her bir kelime karakterlerine ayrıştırılır. Ardından her biri kendi ASCII koduyla değiştirilir. Böylece her bir kelime bir antijen (antigen) olarak düşünülür. Aynı işlem kimliksizleştirilecek verideki kelimelere de uygulanır. Bu verideki kelimeler antikor (antibodies) olarak düşünülür. Tespit etme aşaması mesafe metriğini kullanarak veriden kimlik saptayıcılarının (identifier) tespit edilmesine olanak sağlar. Mesafe metriği her bir antijen ve tüm antikorlar arasında hesaplanır. Depolama adımı bellekteki tanınan antikorlar depolanmaktadır. Doğal bağışıklık sistemi sahibiyle birlikte doğan bir başlangıç değerine sahiptir. Bu çalışmadaki sistemde bu başlangıç değeri kullanıcının yazıtında (inscription) oluşturulur. Böylece yazıt formülünde doldurulan bilgi antijenlere dönüştürülmektedir ve bellekte depolanmaktadır. Sonrasında yeni antikorlar tanınmaya kadar, bellek bu antikorları ekleyerek güncellenir. Sonunda bu antikorlar antijen olarak düşünülen kelimelerin bazı varyasyonlarını temsil eder. Kelimelerin depolandığı bu yol MapReduce prensibine benzemektedir. MapReduce prensibinde kelime anahtarı ve kelimenin konumu da depolama adresini temsil etmektedir. Bu büyük veride eşleme (mapping) olarak adlandırılmaktadır. Yer değiştirme aşamasında kimlik saptayıcıların tespit edilmesi ve depolanması bittikten sonra orijinal veride bellekteki her bir kelimenin karıştırılması ve yer değiştirilmesi ile kimliksizleştirme adımı başlar. Yer değiştirme aşaması şu şekilde çalışır: bir kez tanınan antikor sahibinin belleğinde depolanır. Bu birçok tekrarlama ile devam eder. Her birinde iki antikor rastgele olarak seçilir ve sıralanır. Sonrasında sıralanan kelimedenden iki rastgele bileşen seçilir ve sırası değiştirilir. Ardından sonuç iki bölüm olacak şekilde bir rastgele konum alınır. İlki seçilen ilk antikor ile ikincisi ise ikinci ile yer değiştirilir. Sonuç olarak orijinal metinde yer değiştirme antikorları kimlik saptayıcılarını değiştirmek için çözülür [14].

Tekrardan kimlik saptama kimliksizleştirilmiş belgelerin her bir kelimenin kendi haline yeniden sokulmasıyla orijinal yapısına döndürülmesidir. Bunu yapmanın yolu depolama aşamasında kullanılan matematiksel fonksiyon kadar kolaydır. Kesme işlemi bu aşamada yapılır. Hafızadaki her bir antikor (antibody) her bir kelimenin konumunu belirleyen bir anahtardır. Antikorların konumları belirlendikten sonra sistem; metin dosyalarına eklenecek olan hakiki kelimeleri oluşturmak için bu antikorları çözer (decode) [14].

Büyük verinin gelişimini önleyen en büyük engellerden biri kullanıcıların mahremiyetidir. Bu konu bağlamında birçok gelişmiş araştırma yapılmıştır. Bunlardan biri ise homomorfik şifreleme olarak bilinen kriptografiyle ilgili bir kavramdır. Homomorfik şifreleme, şifre çözmeye ihtiyacı olmadan şifreli veri üzerinde işlemlerin uygulanmasını sağlar. Genetik algoritmalar son

zamanlarda öne sürülen kriptosistemlerinin yeni bir türüdür ve veri madenciliği evrimsel yöntemlerini kullanan değiştirme (substitution) ve yer değiştirme (transposition) gibi klasik kriptografi tekniklerini geliştirmeyi amaçlamaktadır. Bu tip sistemlerin etkinliği IND-CPA (attacks with chosen plaintext) ve IND-CCA (attacks with chosen ciphered text) ile kanıtlanmıştır. Çalışmada TSZ (To, Safavi-Naini ve Zhang) olarak bilinen homomorfik kriptosisteminin etkinliği TSZ ve evrimsel kriptografiyi birleştiren ve iki yöntemin avantajlarını kullanan yeni bir yaklaşım öne sürülerek artırılmıştır [15].

Bulut bilişim; sağlık ve işletme gibi sektörlerdeki çeşitli büyük veri uygulamalarının farklı aşamalarını desteklemek için ölçeklenebilir IT altyapıları sağlamaktadır. Elektronik sağlık kayıtları gibi veri setleri çoğunlukla hassas gizlilik bilgileri içermektedir. Bu bilgilerin bulutta üçüncü şahıslar ile paylaşılması gizlilik endişelerini beraberinde getirir. Veri gizliliği koruması için pratik ve yaygın olarak benimsenen teknik; bir gizlilik modelini sağlayan genelleme (generalization) ile verinin anonimleştirilmesidir. Ancak var olan çoğu gizlilik koruma yaklaşımı yetersizliği veya kötü ölçeklenebilirliği sebebiyle büyük verileri küçük boyutlu veri setlerine dönüştürürler. Zhang ve diğerleri tarafından yapılan çalışmada [16] yakınlık gizlilik ihlallerine karşı büyük veri anonimleştirme için yerel yeniden kodlama problemi (local-recoding problem) incelenmiştir ve bu probleme ölçeklenebilir bir çözüm tanımlamak için çalışılmıştır. Özellikle hassas değerler ve çoklu hassas özelliklerin semantik yakınlığını sağlayan bir yakınlık gizlilik modeli sunulmuştur ve yerel yeniden kodlama problemi bir yakınlık farkındalıklı kümeleme problemi (proximity-aware clustering problem) olarak modellenmiştir. Bu problemleri çözebilmek için t-atalı (ancestors) kümeleme (k-means'e benzer) algoritmasından oluşan ölçeklenebilir iki fazlı kümeleme yaklaşımı ve yakınlık farkındalıklı yığımsal (agglomerative) kümeleme algoritması sunulmuştur. Bulutta veri paralel hesaplama gerçekleştirerek yüksek ölçeklenebilirlik kazanmak için algoritmalar MapReduce ile tasarlanmıştır. Gerçek veri setleri üzerindeki kapsamlı deneyler çalışmadaki yaklaşımın mevcut yaklaşımlar üzerinde yakınlık gizlilik ihlalleri savunma kapasitesini, ölçeklenebilirliği ve yerel yeniden kodlama anonimleştirmesinin zaman verimliliğini artırdığını göstermektedir.

Büyük veri ekonomik büyüme ve teknik yenilik için bilgi madenciliği yapabilmesi sebebiyle son zamanlarda oldukça dikkat çekmektedir ve birçok araştırma çalışması yüksek hacim (volume), velocity (hız) ve çeşitlilik (variety) ("3V" olarak anılır) zorlukları sebebiyle büyük veri işlemeye yönlendirilmiştir. Eğer veri güvenilir (authentic) değilse madencilikten elde edilen yeni bilgi inandırıcı olmayacaktır. Gizlilik iyi bir şekilde sağlanmadıysa insanlar kendi verilerini paylaşmak için isteksiz olabilirler. Çünkü büyük veride güvenlik "gerçeklik (veracity)" olarak adlandırılan yeni bir boyut

olarak incelenmektedir. Lu ve diğerleri tarafından yapılan çalışmada [17] gizlilik açısından büyük verinin yeni zorluklarının kullanılması amaçlanmıştır ve büyük veri çağındaki etkili ve gizlilik korumalı programlama üzerine odaklanılmıştır. İlk olarak büyük veri analitiğinin genel mimarisi biçimlendirilmiştir. Ardından karşılık gelen gizlilik gereksinimleri tanımlanmıştır. Son olarak etkili ve gizlilik koruyucu kosinüs (cosine) benzerlik hesaplama protokolü büyük veri çağında veri madenciliğinin etkinliğine ve gizlilik gereksinimlerine karşılık olarak öne sürülmüştür [17].

Hastane ya da banka gibi bir veri sahibinin gizli olarak saklanan kişiye özgü alanlara ayrılmış veri yığınına sahip olduğunu düşünelim. Bu veri sahibinin verinin bir örneğini araştırmacılarla paylaşmak istediğini varsayalım. Veri sahibi verinin kaynağı olan bireylerin kimliklerinin tekrardan tespit edilemeyecek şekilde bilimsel teminatlı özel verinin örneğini nasıl paylaşmalıdır? Sweeney tarafından yapılan çalışmada [18] öne sürülen çözüm k-anonimlik olarak adlandırılan biçimsel bir koruma modeli ve dağıtım için bir dizi ilkeyi içermektedir. Bir yayın (release) o yayındaki her bir kişinin bilgisi, yine o yayında bilgisi olan en az k-1 birey tarafından ayırt edilemiyorsa k-anonimlik korumasını temin eder. Ayrıca çalışmada k-anonimliğe bağlı kalan yayınlar üzerinde gerçekleştirilebilen tekrardan kimlik saptama atakları incelenmiştir. k-anonimlik koruma modeli gerçek dünya sistemlerine dayandığı için önemlidir.

Birçok kurum halk sağlığı ve nüfus araştırması gibi amaçlarla mikro verilerini yayınlamaktadır. İsim, sosyal güvenlik numarası gibi bireyleri açık bir şekilde belirleyen özellikler genel olarak silinseler de bu veritabanları bazen posta kodu, cinsiyet ve doğum tarihi gibi özellikler üzerinden anonim kalması gereken bireyleri tekrardan belirlemek için diğer açık veritabanları ile birleştirilirler. Birleştirme atakları (joining attacks) Internet üzerinden birbirini tamamlayan diğer veritabanlarının olmasıyla daha kolay gerçekleştirilir [19].

K-anonimleştirme (anonymization) yayınlanan mikro verinin parçalarını genelleyerek ve/veya gizli tutarak birleştirme ataklarını önleyen bir tekniktir. Böylece k boyutundaki bir gruptaki herhangi bir birey eşsiz olarak ayırt edilemez. LeFevre ve diğerleri çalışmasında [19] k-anonimleştirme modelini gerçeklemek için tam alan genelleme (full-domain generalization) olarak adlandırılan uygulanabilir bir çerçeve sunmuştur. Minimum tam alan genellemeyi gerçekleştirmek için bir algoritma dizisi sunulmuştur ve bu algoritmaların iki gerçek veritabanı üzerinde önceki algoritmalarından büyüklük sırasına göre daha hızlı olduğu gösterilmiştir.

Veri kimliksizleştirme araştırma amaçlı verinin yayınlanması ihtiyaçlarını ve bireylerin gizlilik isteklerini uzlaştırmaktadır. Bayardo ve Agrawal çalışmasında [20] k-anonimleştirme olarak bilinen güçlü kimliksizleştirme yöntemi için bir optimizasyon algoritması öne sürmüştür ve değerlendirmelerini yapmıştır. k-anonimleştirilmiş bir veri setinde her bir kayıt en az diğer k-1 kayıttan ayırt

edilemez olmalıdır. Optimize edilmiş k-anonimliğin basit kısıtları NP-hard olsa bile önemli hesaplamalı zorluklara neden olmaktadır. Gerçek nüfus sayımı verileri üzerinde yapılan deneylerde öne sürülen algoritmanın iki temsili maliyet ölçümü ve geniş bir k aralığı altında en iyi k-anonimleştirmelerini bulabildiği gösterilmiştir. Ayrıca algoritmanın giriş verisi veya giriş parametrelerinin makul sürede en iyi çözümü bulmayı engellediği durumlarda da iyi anonimleştirme ürettiği gösterilmiştir. Son olarak algoritma; anonimleştirme kalitesi ve performans üzerinde farklı kodlama yaklaşımlarının ve problem varyasyonlarının etkilerini araştırmak için kullanılmıştır. Yazarlara göre bu çalışma problemin genel bir modeli altında çözülmesi zor alan bir veri setinin en iyi k-anonimleştirmesini kanıtlayan ilk sonuçtur.

K-anonimlik mikro veri yayınlamada gizliliği korumak için öne sürülen bir mekanizmadır ve k-anonimliği sağlamak için çok sayıda kodlama modeli düşünülmüştür. LeFevre ve diğerleri tarafından yapılan çalışmada [21] önceki çalışmalarda (tek-boyutlu) görülmeyen ek esneklik derecesi sağlayan yeni bir çok boyutlu (multidimensional) model öne sürülmüştür. Bu esneklik genel amaçlı ölçü birimleri ve daha belirli sorgu cevaplanabilirlik kavramları ile ölçüldüğü gibi daha yüksek kalitede anonimleştirmeye yol açmaktadır. Optimal çok boyutlu anonimleştirme NP-hard'dır. Fakat çalışmada basit bir greedy yaklaşım algoritması öne sürülmüş ve deneysel sonuçlar bu algoritmanın iki tek boyutlu model için ayrıntılı (exhaustive) optimal algoritmadan daha makul anonimleştirme yaptığını göstermektedir.

Sweeney çalışmasında [22] k-anonimliği (anonymity) gerçekleştirmek için genelleme (generalization) ve gizlemeyi (suppression) birleştiren biçimsel bir sunum ortaya koymuştur. Genelleme bir değer daha az belirli fakat anlamsal olarak tutarlı bir değer ile değiştirilmesi (veya yeniden kodlanması) ile ilgilidir. Gizleme ise bir değer hiçbir biçimde yayınlanmamasıdır. MinGen (Minimal Generalization Algorithm, Minimum Genelleme Algoritması) teorik bir algoritmadır ve minimum bozulmalı k-anonimlik koruması sağlamak için bu iki tekniği birleştirmektedir.

Gizlilik; sosyal bilim araştırması ya da iş analizi için sosyal ağ verilerinin yayınlanmasında veya paylaşılmasında en önemli sorunlardan biridir. Son zamanlarda araştırmacılar yapı bilgisi yoluyla düğüm kimlik saptamasını engellemek için k-anonimliğe gizlilik modelleri geliştirmişlerdir. Ancak bu gizlilik modelleri uygulansa bile bir düğüm grubu yaygın bir şekilde aynı hassas etiketleri (özellikleri) paylaşıyorsa saldırgan, birinin gizli bilgisine ulaşabilir. Diğer bir deyişle burada etiket düğüm ilişkisi saf yapı anonimleştirme yöntemleri ile iyi korunmamaktadır. Ek olarak uç düzenleme ya da düğüm kümelemeye dayanan var olan yaklaşımlar önemli ölçüde anahtar grafik özelliklerini değiştirebilir. Yuan ve diğerleri çalışmasında [23] bireylerin hassas etiketlerinin yanında yapısal bilgi korumasını göz önünde bulunduran bir k-derece-1-çeşitlilik anonimlik modeli (k-degree-1-diversity anonymity model) tanımlamıştır. k-derece-1-

çeşitliliğin gereksinimlerini sağlamak için orijinal grafdan yeni bir graf oluşturmak için gürültü (noise) düğüm ekleme algoritması tasarlanmıştır. Burada orijinal graf üzerinde en az bozulma kısıtlaması vardır. Çalışmada gerçekleştirilen deney sonuçları gürültü düğüm ekleme algoritmasının yalnızca uç düzenleme kullanan çalışmalardan daha iyi sonuçlar verdiğini göstermektedir.

iRODS (Integrated Rule-Oriented Data, Entegre Kural-Odaklı Veri) isimli yeni bir teknik büyük veride güvenliği ve gizliliği sağlamak için öne sürülmüştür [1, 24]. iRODS'un en önemli teknik özellikleri şunlardır: birleşmiş veri sistemleri (grids) veya akıllı bulutlar, dağıtık kural motoru, iCAT metadata kataloğu, ortak erişim sağlayan depolama erişim katmanı, grafiksel kullanıcı arayüzü ve komut-satırı-tabanlı istemcilerin

zengin kombinasyonu ve iRODS veri sistemi ile etkileşimi sağlayan API'ler. iRODS birçok veri yönetimi uygulamasında kullanılmıştır ve dünya çapında çok sayıda kuruluş tarafından benimsenmiştir. iRODS teknolojisinin kural tabanlı ve büyük ölçekli veri yönetimindeki çeşitli zorlukların üstesinden gelmek için nasıl uygulandığıyla alakalı birçok yayın vardır [1, 25-28]. iRODS teknolojisi veriyi korumak ve gizliliği garanti etmek için olan yaygın yaklaşımlar üzerinde gelişme sağlamaktadır. Bu gelişmeler şunlardır: güvenlik kontrollerinin kapsamlı seti, veri erişiminin gelişmiş kontrolü ve metadata üzerinden kullanımı, depolama sanallaştırması ve veri güvenliği yaşam döngüsü ve kalıcı tanımlayıcılar (persistent identifiers) [1]. Tablo 1'de büyük veri güvenliği için geliştirilen yöntemler özetlenmektedir.

Tablo 1. Büyük veri güvenliği için geliştirilen yöntemler (atıf sayısına ve kronolojiye göre)
(The methods developed for big data security (in terms of the number of citation and chronology))

Yayın İsmi	Yazar	Yıl	Yöntem	Atıf Sayısı
k-anonymity: a model for protecting privacy	Sweeney [18]	2002	<i>k</i> -anonimlik olarak adlandırılan biçimsel bir koruma modeli öne sürülmüştür.	4179
Achieving k-anonymity privacy protection using generalization and suppression	Sweeney [22]	2002	<i>k</i> -anonimliği gerçekleştirmek için genelleme (generalization) ve gizlemeyi (suppression) birleştiren biçimsel bir sunum ortaya konulmuştur. MinGen (Minimal Generalization Algorithm) teorik bir algoritmadır ve minimum bozulmalı <i>k</i> -anonimlik koruması sağlamak için bu iki tekniği birleştirmektedir.	1601
Incognito: Efficient Full-Domain K-Anonymity	LeFevre ve diğ. [19]	2005	<i>k</i> -anonimleştirme modelini gerçeklemek için tam alan genelleme olarak adlandırılan uygulanabilir bir çerçeve sunulmuştur ve iki gerçek veritabanı üzerinde önceki algoritmalarından hızlı olduğu gösterilmiştir.	1168
Data privacy through optimal k-anonymization	Bayardo ve Agrawal [20]	2005	<i>k</i> -anonimleştirme olarak bilinen güçlü kimliksizleştirme yöntemi için bir optimizasyon algoritması öne sürülmüş ve değerlendirilmiştir. Bu çalışma problemin genel bir modeli altında çözülmesi zor alan bir veri setinin en iyi <i>k</i> -anonimleştirmesini kanıtlayan ilk sonuçtur.	1125
Mondrian Multidimensional K-Anonymity	LeFevre ve diğ. [21]	2006	Önceki çalışmalarda (tek-boyutlu) görülmeyen ek esneklik derecesi sağlayan yeni bir çok boyutlu (multidimensional) model öne sürülmüştür.	1003
Toward efficient and privacy-preserving	Lu ve diğ. [17]	2014	Gizlilik açısından büyük verinin yeni zorluklarının kullanılması amaçlanmıştır ve büyük veri çağındaki etkili ve	86

computing in big data era			gizlilik korumalı programlama üzerine odaklanılmıştır.	
Protecting Sensitive Labels in Social Network Data Anonymization	Yuan ve diğ. [23]	2013	Bireylerin hassas etiketlerinin yanında yapısal bilgi korumasını göz önünde bulunduran bir <i>k</i> -derece- <i>l</i> -çeşitlilik anonimlik modeli tanımlanmıştır.	55
Proximity-Aware Local-Recoding Anonymization with MapReduce for Scalable Big Data Privacy Preservation in Cloud	Zhang ve diğ. [16]	2015	Yakınlık gizlilik ihlallerine karşı büyük veri anonimleştirme için yerel yeniden kodlama problemi (local-recoding problem) incelenmiştir ve bu probleme ölçeklenebilir bir çözüm tanımlamak için çalışılmıştır.	22
Big Data security and privacy: A review	Matturdi ve diğ. [1]	2015	Büyük veri güvenliği ve mahremiyeti hakkında genel bilgi verilmektedir.	18
A Multilayer Evolutionary Homomorphic Encryption Approach for Privacy Preserving over Big Data	Rahmani ve diğ. [15]	2014	TSZ (To, Safavi-Naini ve Zhang) olarak bilinen homomorfik kriptosisteminin etkinliği TSZ ve evrimsel kriptografiyi birleştiren ve iki yöntemin avantajlarını kullanan yeni bir yaklaşım öne sürülerek artırılmıştır.	7
An approach towards big data — A review	Gupta ve Tyagi [11]	2015	Büyük verinin işlevsel güvenliği, siber güvenlikte büyük veri analizleri ve tehdit tespiti için güvenlik analizleri gibi konularda genel bilgi verilmektedir.	5
Big Data: Big Promises for Information Security	Alguliyev ve Imamverdiyev [9]	2014	Büyük veri güvenliği ve siber güvenlik için olan büyük veri zorlukları hakkında genel bilgi verilmektedir.	5
De-identification of Textual Data Using Immune System for Privacy Preserving in Big Data	Rahmani ve diğ. [14]	2015	Özel bir bağışıklık sistemi kullanarak metin verilerini kimliksizleştiren yeni bir model ortaya konulmuştur. Bu yaklaşımda kimliksizleştirme ve tekrardan kimlik saptama olmak üzere iki ana aşama vardır.	2

Bu çalışmada büyük veri güvenliği için geliştirilen yöntemler kısaca açıklanmıştır ve bu alanda çalışacak olan yeni araştırmacılara yol göstermek hedeflenmiştir. İncelenen makalelere bakılarak bir değerlendirme yapılmak istenildiğinde büyük veri üzerindeki güvenlik ve kişi mahremiyeti mekanizmalarının henüz olgunlaşmamış olduğu ve sürekli olarak geliştiği görülmektedir.

6. SONUÇ (CONCLUSION)

Bu çalışmada büyük veride kişi mahremiyetinin korunması ve büyük veri güvenliği ile ilgili bugüne kadar yapılmış olan bilimsel çalışmalar özetlenmiştir. Ayrıca büyük veri sistemlerini korumak için gerekli olan araçlardan ve bu sistemlerin güvenliğinin sağlanmasında

karşılaşılan zorluklardan bahsedilmiştir. Buna ek olarak büyük veri kaynaklarının neler olduğu açıklanmıştır.

Günümüzde medya paylaşımları, sosyal ağ gibi farklı kaynaklardan günlük olarak birçok veri üretilmektedir ve veri miktarı hızlı bir şekilde artmaktadır. Bu da büyük veri kavramını ortaya çıkarmıştır ve kişi mahremiyeti noktasında farklı birçok zorluğa sebep olmuştur. Bu alanda yapılan çalışmaların giderek artması gelecekte bu konunun daha da önem kazanacağını göstermektedir.

TEŞEKKÜR (ACKNOWLEDGEMENT)

Bu çalışma İstanbul Üniversitesi Fen Bilimleri Enstitüsü Bilgisayar Mühendisliği Anabilim Dalı'na bağlı olarak yürütülen "Büyük Veride Etkin Gizlilik Koruması için

Yazılım Tasarımı” adlı doktora tezinin bir bölümüdür. Ayrıca İstanbul Ticaret Üniversitesi Bilişim Teknolojileri Uygulama ve Araştırma Merkezi tarafından desteklenmektedir.

KAYNAKLAR (REFERENCES)

- [1] B. Matturdi, X. Zhou, S. Li, F. Lin, “Big Data security and privacy: A review”, *China Communications*, 11(14), 135-145, April 2015.
- [2] N. Perlroth, J. Larson, S. Shane, **NSA able to foil basic safeguards of privacy on web**, The New York Times, 6, 5 September 2013.
- [3] P. Ohm, “Broken promises of privacy: Responding to the surprising failure of anonymization”, *UCLA Law Review*, 57, 1701, 2010.
- [4] K. LeFevre, D. J. DeWitt, R. Ramakrishnan, “Mondrian multidimensional k-anonymity”, **Proceedings of the 22nd International Conference on Data Engineering**, IEEE, 25-25, 03-07 April 2006.
- [5] L. Lü, T. Zhou, “Link prediction in weighted networks: The role of weak ties”, *EPL (Europhysics Letters)*, 89(1), 18001, January 2010.
- [6] A. Clauset, C. Moore, M. E. J. Newman, “Hierarchical structure and the prediction of missing links in networks”, *Nature*, 453, 98-101, May 2008.
- [7] D. Yin, L. Hong, X. Xiong, “Link formation analysis in microblogs”, **Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval**, 1235-1236, 2011.
- [8] R. N. Lichtenwalter, J. T. Lussier, N. V. Chawla, “New perspectives and methods in link prediction”, **Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining**, 243-252, 2010.
- [9] R. Alguliyev, Y. Imamverdiyev, “Big Data: Big Promises for Information Security”, **2014 IEEE 8th International Conference on Application of Information and Communication Technologies (AICT)**, 1-4, Astana, 15-17 October 2014.
- [10] **Cloud Security Alliance (CSA): Expanded Top Ten Big Data Security and Privacy Challenges**, April 2013.
- [11] P. Gupta, N. Tyagi, “An approach towards big data — A review”, **2015 International Conference on Computing, Communication & Automation (ICCCA)**, 118-123, Noida, 15-16 May 2015.
- [12] Internet: Securing Big Data: Security Recommendations for Hadoop and NoSQL Environments, https://securisis.com/assets/library/reports/SecuringBigData_FINAL.pdf, 22.09.2016.
- [13] T. Mahmood, U. Afzal, “Security Analytics: Big Data Analytics for Cybersecurity”, **2013 2nd National Conference on Information Assurance (NCIA)**, 2013.
- [14] A. Rahmani, A. Amine, M. R. Hamou, “De-identification of Textual Data Using Immune System for Privacy Preserving in Big Data”, **2015 IEEE International Conference on Computational Intelligence & Communication Technology (CICT)**, 112-116, Ghaziabad, 13-14 February 2015.
- [15] A. Rahmani, A. Amine, R. H. Mohamed, “A Multilayer Evolutionary Homomorphic Encryption Approach for Privacy Preserving over Big Data”, **2014 International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery (CyberC)**, 19-26, Shanghai, 13-15 October 2014.
- [16] X. Zhang, W. Dou, J. Pei, S. Nepal, C. Yang, C. Liu, J. Chen, “Proximity-Aware Local-Recoding Anonymization with MapReduce for Scalable Big Data Privacy Preservation in Cloud”, *IEEE Transactions on Computers*, 64(8), 2293-2307, 2015.
- [17] R. Lu, H. Zhu, X. Liu, J. K. Liu, J. Shao, “Toward efficient and privacy-preserving computing in big data era”, *IEEE Network*, 28(4), 46-50, 2014.
- [18] L. Sweeney, “k-anonymity: a model for protecting privacy”, *International Journal on Uncertainty, Fuzziness and Knowledge-based Systems*, 10(5), 557-570, 2002.
- [19] K. LeFevre, D. J. DeWitt, R. Ramakrishnan, “Incognito: Efficient Full-Domain K-Anonymity”, **ACM SIGMOD international conference on Management of data**, 49-60, Baltimore, Maryland, USA, 14-16 June 2005.
- [20] R. J. Bayardo, R. Agrawal, “Data privacy through optimal k-anonymization”, **21st International Conference on Data Engineering**, IEEE, 217-228, 5-8 April 2005.
- [21] K. LeFevre, D. J. DeWitt, R. Ramakrishnan, “Mondrian Multidimensional K-Anonymity”, **22nd International Conference on Data Engineering**, IEEE, 25, 03-07 April 2006.
- [22] L. Sweeney, “Achieving k-anonymity privacy protection using generalization and suppression”, *International Journal on Uncertainty, Fuzziness and Knowledge-based Systems*, 10(5), 571-588, 2002.
- [23] M. Yuan, L. Chen, P. S. Yu, T. Yu, “Protecting Sensitive Labels in Social Network Data Anonymization”, *IEEE Transactions on Knowledge and Data Engineering*, 25(3), 633-647, 2013.
- [24] M. Jensen, “Challenges of privacy protection in big data analytics”, **2013 IEEE International Congress on Big Data (BigData Congress)**, 235-238, 2013.
- [25] A. Rajasekar, R. Moore, C. Hou, “iRODS Primer: integrated rule-oriented data system”, *Synthesis Lectures on Information Concepts, Retrieval, and Services*, 2(1), 1-143, 2010.
- [26] A. Rajasekar, R. Moore, M. Wan, “Applying rules as policies for large-scale data sharing”, **2010 International Conference on Intelligent Systems, Modelling and Simulation (ISMS)**, IEEE, 322-327, 2010.
- [27] I. Barg, D. Scott, E. Timmermann, “NOAO E2E integrated data cache initiative using iRODS”, **XX. ASP Conference Proceedings Astronomical Data Analysis Software and Systems**, 442, 497-500, 2011.
- [28] J. L. Schnase, W. P. Webster, L. A. Parnell, “The NASA Center for Climate Simulation Data Management System”, **2011 IEEE 27th Symposium on Mass Storage Systems and Technologies (MSST)**, 1-6, 2011.

Vardiya Çizelgeleme Problemi ve Bir Örnek Uygulama

Emre VARLI, Tamer EREN

Endüstri Mühendisliği, Kırıkkale Üniversitesi, Kırıkkale, Türkiye
ef.varli@gmail.com, tamereren@gmail.com
 (Geliş/Received: 12.10.2016; Kabul/Accepted: 12.01.2017)
 DOI: 10.17671/gazibtd.309302

Özet— Günümüzde birçok işletme daha iyi bir hizmet veya üretim yapmak için birbirleriyle rekabet içerisinde. Bu nedenle her işletme varlıklarını sürdürebilmek ve verimliliğini artırmak için çalıştırdığı personelin istek ve tercihlerini dikkate alarak müşterilerine daha kaliteli ürün veya hizmet sağlayabilirler. Genellikle personel memnuniyeti için kullanılan personel çizelgeleme, son yıllarda popüler bir problem haline gelmiştir. Uygulama alanı çok geniş bir yelpazeye sahip olduğundan çok yoğun bir şekilde kullanılmaktadır. Bu çalışmada, bir fabrikada çalışan şefleri ilk olarak Analitik Hiyerarşi Prosesi (AHP) tekniğinden faydalanarak belirli kriterler yardımıyla kıdem seviyelerine ayırdıktan sonra fabrikanın her bir vardiya için ihtiyaç duyduğu işgücünü karşılamak, planlanan dönem boyunca şeflere verilen izin günlerini düzenlemek, her bir şefin toplam işgücünü dengeli ve adaletli bir şekilde dağıtmak ve en önemlisi kıdem seviyelerini kullanarak niteliği çok olan şefler ile niteliği az olan şefleri aynı vardiyaya olabildiğince atanması gibi problemlerin çözümü için bir model geliştirilmeye çalışılmıştır.

Anahtar Kelimeler— Kıdem Seviyeleri, Analitik Hiyerarşi Prosesi, Vardiya Çizelgeleme, Hedef Programlama

Shift Scheduling Problems and A Case Study

Abstract— Many businesses nowadays to make a better service or production are in competition with each other. Therefore, to maintain their existence taking into consideration the wishes and the preferences of each business and staff to improve the efficiency of the run can provide a higher quality product or service to customers. Staff scheduling generally used for staff satisfaction so the problem in recent years has become a popular. Application area of a very broad range is used very intensively. In this study, a factory employee chefs firstly Analytic Hierarchy Process (AHP) after separating the seniority level with the help of specific criteria, taking advantage of the technique to meet the labor force needed for each shift of the factory, during the planning period to organize the day allowed the chef, balanced and equitable total workforce of each chief a way to distribute and most importantly the skills of the chefs with less seniority levels using the very skills of the chief has attempted to develop a model for the solution of problems such as assigning the same shift.

Keywords— Seniority Level, Analytical Hierarchy Process, Shift Scheduling, Goal Programming

1. GİRİŞ (INTRODUCTION)

Son yıllarda işletmeler, topluma daha iyi bir hizmet vermek adına birbirleriyle rekabet içerisinde. Bu rekabet ortamında ayakta kalabilmek ve insanların beklentilerini karşılayabilmek için herhangi bir işletmede çalışan personelin verimliliğinin ve kalitesinin arttırmasına yönelik çalışmalar yapılması gereklidir. Personel memnuniyetinin arttırılmasında personel çizelgelemenin önemi çok büyüktür. Çünkü personel çizelgeleme, çalışan personelin ihtiyaç ve tercihleri doğrultusunda bir model oluşturarak hem çalışan personelin hem de ilgili işletmenin planlı ve sistemli bir şekilde üretim veya hizmet yapabilesini sağlamaktadır.

Personel çizelgelemenin bir alt problemi olan vardiya çizelgeleme problemi, geçmişten günümüze kadar yapılan çalışmalar incelendiğinde en çok üzerinde çalışılan bir problem olarak karşımıza çıkmaktadır. Vardiya çizelgeleme problemi sanayilerde, işletmelerde, hastanelerde ve buna benzer birçok yerde karşımıza çıkan önemli bir problemdir. Bu problemin çözümünde, çalışan personelin vardiyalara adaletli ve dengeli bir şekilde dağıtarak onların insanlara daha iyi bir hizmet vermesi amaçlanmaktadır. Bunlara ek olarak vardiya çizelgeleme problemi, çalışan personelin yanı sıra işçi ücretleri, fazla mesailer, mola saatleri gibi konularda iyileştirmeler yaparak kurum ve kuruluşların gelirlerinde de artış sağlamaktadır.

Çalışan personele uygun bir planlama yapmak için personel tercihlerini, niteliklerini ve buna benzer birçok kriter hesaba katılmalıdır. Böylelikle el ile yapılan çizelgelerin pek bir öneminin kalmaması ile birlikte analitik yöntemlerin kullanılması gerekmektedir. Ek olarak personele uygun planın yapılması verimliliği artıracak ve müşterilerin istekleri daha çabuk karşılanacaktır [1].

Hedef programlama son yıllarda yaygın olarak kullanılan çok ölçütlü karar verme tekniklerinden biridir. Hedef programlamada amaç, birbirleriyle çakışan birçok hedefin arasından istenilen sonuçları elde etme çabasıdır. Hedef programlama matematiksel modellemede kullanılan tekniklerdeki gibi optimal sonucu bulmayla ilgilenmez. Hedef programlamada istenilen hedeflerdeki sapma değişkenlerini en küçükleyerek hedefleri olabildiğince gerçekleştirme arzusudur.

Çalışmada kullanılan başka bir yöntemde Analitik Hiyerarşi Prosesi (AHP) yöntemidir. İlk defa Satty tarafından kullanılan çok ölçütlü karar verme yöntemlerinden biridir. Bu yöntemde karar hiyerarşisi tanımlandığında kararları etkileyen faktörlerin, karar aşamasında yüzde dağılımlarını veren bir karar verme yöntemi olarak açıklanabilir.

Bu çalışmada bir fabrikada çalışan şefleri çok kriterli karar verme yöntemlerinden biri olan Analitik Hiyerarşi Prosesi (AHP) yöntemi ile kıdemlerine göre sınıflandırarak ve daha sonra 28 günlük çalışma periyodunda fabrikanın ihtiyaç duyduğu iş gücünü karşılamak, şefleri dengeli bir şekilde vardiyalara atamak, şeflere verilen izin günlerini düzenlemek ve en önemlisi kıdem seviyelerine göre ayrılan şeflerin her vardiyaya atanırken genellikle en niteliklisi ile en niteliksizi aynı vardiyaya atayarak fabrikanın çalışma temposunu artırmak ve kıdemi daha çok olan şefler ile daha az olan şefler arasında düzeni sağlamak amaçlanmıştır. Hedef programlama ile kıdemli-kıdemsiz ataması, aylık işgücü dağılımının eşit olması, modelin daha sistemli ve gerçeğe uygun olması için kullanılan sapma değişkenleri minimize edilerek model kurulmuştur.

Yapılan bu çalışmada, ikinci bölümde vardiya çizelgeleme, üçüncü bölümde hedef programlama, dördüncü bölümde analitik hiyerarşi proses (AHP), beşinci bölümde literatür araştırması, altıncı bölümde yapılan uygulama ve son olarak yedinci bölümde de yapılan uygulamanın sonuçları anlatılmıştır.

2. VARDİYA ÇİZELGELEME (SHIFT SCHEDULING)

Vardiya çizelgeleme, personel çizelgeleme problemleri arasında en çok ele alınan problem çeşidi olarak gösterilebilir. Yapılan çalışmalar incelendiğinde vardiya çizelgeleme problemi neredeyse bütün çalışmalarda ele alınmıştır. Fabrikalar veya işletmeler vardiya çizelgeleme ile personel için bir çalışma planı yaparak personel memnuniyeti, verimliliği, işletme giderleri ve zaman gibi konularda iyileştirmeler yapabilirler. Ek olarak çalışılan

yerin kendine özgü benimsediği kuralları modele ekleyerek gerçeğe uygun ve çalışma prensiplerini yerine getirerek farklı çizelgelemeler yapılabilir.

Personel çizelgeleme problemleri ilgili işyerinin kuralları doğrultusunda, her çalışanın çalışma tercihlerini dikkate alarak çeşitli hizmet türlerini karşılamak adına nitelikli personeller tahsis edilmesi gereklidir. Bu problemler bazen çözülmesi güç, kısıtlı optimizasyon problemleridir [2].

Genel olarak bu konularda yapılan ilk çalışma Edie ve Dantzing tarafından 1950 yıllarındadır. Dantzing ele aldığı çalışmada araç kabin operatörleri için bir çizelgeleme yapmış ve çözüm olarak liner programlamayı kullanmıştır. Çalışmada gerekli iş gücü miktarının olması gerektiği kadar ya da bundan daha fazla olması koşulu ile maliyeti en küçüklemeyi istemiştir [3-4].

Vardiya çizelgelemenin kullanıldığı alanlardan söz edilecek olursa kamu hizmetleri, fabrika, işletme, turizm ve otelcilik, hastane, demiryolu, havayolu, güvenlik hizmetleri, çağrı merkezleri, perakende ve askeri sistemler alanlarında vardiya çizelgeleme yoğun bir şekilde kullanılmaktadır. Sektör olarak bu kadar geniş bir yelpazeye sahip olması vardiya çizelgelemenin yaşantımızda ne kadar önemli bir yere sahip olduğunu da göstermektedir.

Vardiya çizelgeleme probleminin çözümü için matematiksel modellemelerden, sezgisel algoritmalarından veya her ikisini birden kullanarak model oluşturulur. Matematiksel modeller: doğrusal programlama, tam sayılı programlama, hedef programlama, dinamik programlama ve sezgisel algoritmalar: genetik algoritma, tabu arama yöntemi, stokastik programlama bu yöntemlerden birkaçı olarak gösterilebilir.

3. HEDEF PROGRAMLAMA (GOAL PROGRAMMING)

Hedef programlama çok kriterli karar verme modellerinin bir türüdür. Modeli oluşturulurken olmazsa olmaz kısıtlara ek olarak hedeflenen kısıtlara gevşek değişkenler eklenerek modelin kısıtları yazılır. Buradaki amaç oluşturulan hedef kısıtlarındaki gevşek değişkenlerin minimize edilmesi işlemidir. Diğer modellemelerde tek bir amaç fonksiyonu kullanılırken hedef programlama ile oluşturulan modellerde birden çok amaç aynı anda veya sırasıyla kullanılır yani hedef programlama birden fazla amacı birer kısıt haline dönüştürerek kısıtların önem derecelerine göre sıralayarak veya her birini ağırlıklandırarak bu kısıtların sapmalarını amaç fonksiyonunda en küçükleme ile hedefe ulaşılmasını sağlar.

Hedef programlama her amacın verilen hedeflere mümkün olduğunca ulaşmasını amaçlar ve hedefteki sapmaları minimize etmektedir [5].

Hedef programlama ile ilgili ilk olarak 1955' te Charnes ve arkadaşları çalışmışlardır daha sonra 1961' de Charnes

ve Cooper bu modellemeyi geliştirmek için uğraşmışlardır [6-7].

Hedef programlama matematiksel gösterimi aşağıdaki gibidir [8].

$$\text{Minimize } Z = \sum_{i=1}^k (d_i^+ + d_i^-)$$

$$\sum_{j=1}^n a_{ij} x_j + d_i^+ + d_i^- = b_i$$

$$d_i^+ + d_i^- = 0$$

$$x_j, d_i^+, d_i^- \geq 0 \quad i=1\dots k \quad j=1\dots n$$

Değişkenler

x_j : j . Karar değişkeni

a_{ij} : i . hedefin j . karar değişkeni katsayısı

b_i : i . hedef için ulaşılmak istenen değer

d_i^+ : i . hedefin pozitif sapma değişkeni

d_i^- : i . hedefin negatif sapma değişkeni

Bu modellemeyi kullanarak personel çizelgeleme ile ilgili literatürde pek çok çalışma mevcuttur. Yapılan çalışmalarda genellikle vardiya çizelgeleme problemi ele alınmıştır ve buna ek olarak ekip çizelgeleme, izin günlerini çizelgeleme, iş gücü çizelgeleme gibi problemlerin çözümünde de hedef programlama metodu kullanılmıştır. Yapılan bazı çalışmalar şunlardır: Aly ve Louly [9] telekomünikasyon merkezinde personel çizelgeleme problemi, Azaiez ve Sharif [10] hedef Programlama ile hemşire çizelgeleme, Bektur ve Hasgül [11] kıdem seviyelerine göre iş gücü çizelgeleme, Chu [12] Hong Kong Uluslararası hava alanında çalışan ekibin çizelgenmesi, Elomri vd. [13] bir hastanede çalışan stajyer doktorlar için hedef programlama, Hung-Tso vd. [14] ekip atama problemi, Sulak ve Bayhan [15] tur çizelgeleme problemi, Topaloğlu [16] çalışanların çalışma tercihlerini dikkate alarak tur çizelgeleme problemi, hedef programlama ile yapılan çalışmalara örnek olarak verilebilir.

4. ANALİTİK HİYERARŞİ PROSESİ (ANALYTICAL HIERARCHY PROCESS)

Analitik Hiyerarşi Proses (AHP) karar verme problemine göre belirlenen kriteri kullanarak bu kriterlerin kendi aralarındaki önem ağırlıklarını elde etmede ve kriterleri etkileyen faktörler arasında bir hiyerarşi oluşturmada kullanılan bir yöntem olarak tanımlanabilir. AHP yöntemiyle ilgili ilk çalışma Myers ve Alpert tarafından 1968 yılında yapılmıştır. Saaty, 1977 yılında bu yöntemi bir model haline getirerek karar verme problemleri için daha kullanışlı hale getirmiştir [17]. AHP yöntemini kullanmak için aşağıdaki adımlar dikkate alınmalıdır.

4.1. Karar Verme Probleminin Tanımlanması (Defining Decision Making Problem)

Bu bölümde yapılması gereken iki adım vardır. Birinci adımda kriterler, ikinci adımda ise kriterleri etkileyen faktörler belirlenir. Kriterlerin sayısı a ile faktörlerin sayısı b ile sembolize edilmiştir. Problem için kullanılan kriterlerin ve faktörlerin çok iyi tanımlanması gerekmektedir çünkü birbirleri ile karşılaştırma yapılırken doğru ve tutarlı bir sonuç vermesi açısından önemlidir.

4.2. Kriterler ve Faktörler Arası Karşılaştırma Matrislerinin Oluşturulması (Creating Criteria and Factor Comparison Matrices)

Kriterler ve faktörler sayıları kadar için aşağıda verilen matris oluşturulur ve $a \times a$ boyutta karesel bir matristir. Kullanılan her bir kriter veya faktör kendisi ile karşılaştırılmayacağından dolayı matristeki köşegenler 1 değerini alır.

$$M = \begin{bmatrix} m_{11} & m_{12} & \dots & m_{1a} \\ m_{21} & m_{22} & \dots & m_{2a} \\ \vdots & \vdots & \ddots & \vdots \\ m_{a1} & m_{a2} & \dots & m_{aa} \end{bmatrix}$$

Kriterlerin ve faktörlerin karşılaştırmaları, önem değerleri dikkate alınarak puanlandırma yapılır. Puanlandırma işlemi Saaty'nin 1990'da kullandığı önem skalasına göre yapılmıştır [18].

Karşılaştırmalar yapılırken matristeki köşegen değerlerinin üstünde kalan yerlere puanlar verilir. Köşegen değerlerinin altında kalan yerlerin alacağı puanlarda (4.1)'deki formül yardımıyla elde edilir.

$$m_{ji} = \frac{1}{m_{ij}} \quad (4.1)$$

4.3. Kriterlerin ve Faktörlerin Yüzde Önem Dağılımlarının Belirlenmesi (Determination of Percentage Distribution of Importance Factors and Criterias)

Kriterlerin ve faktörlerin yüzde önem dağılımlarının bulunabilmesi için karşılaştırma matrislerinden yararlanılır. Tüm matrisler için aynı işlemler yapılır. Bu işlemler sırasıyla şu şekildedir. Her bir matrisin sütun vektörleri toplanır daha sonra sütun vektörlerinin toplamı matristeki her bir değere bölünerek başka bir matris elde edilir. Elde edilen matrise N matrisi veya Normalize Matris denilmiştir.

Her bir sütun vektörünün hesaplanmasında aşağıdaki (4.2) formülünden yararlanılır.

$$n_{ij} = \frac{m_{ij}}{\sum_{i=1}^a m_{ij}} \quad (4.2)$$

$$N = \begin{bmatrix} n_{11} & n_{12} & \dots & n_{1a} \\ n_{21} & n_{22} & \dots & n_{2a} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ n_{a1} & n_{a2} & \dots & n_{aa} \end{bmatrix}$$

Bir sonraki adım olan tutarlılık ölçülerinin bulunabilmesi için öncelik vektörlerini elde etmemiz gereklidir. Öncelik vektörü V matrisi ile gösterilmiştir. Öncelik matrislerinin her biri aşağıda verilmiş olan formül (4.3) ile elde edilir.

$$v_{ij} = \frac{\sum_{j=1}^a n_{ij}}{a} \quad (4.3)$$

V öncelik vektörü aşağıda gösterilmiştir.

$$V = \begin{bmatrix} v_1 \\ v_2 \\ \cdot \\ \cdot \\ v_a \end{bmatrix}$$

4.4. Kriterler ve Faktör Kıyaslamalarındaki Tutarlılık Ölçüsü (Consistency Measures of Criteria and Factor Benchmarks)

AHP yöntemi çok sistematik bir yöntem olarak gözüke de yapılan her karşılaştırmanın tutarlı olması istenmektedir. Bunun içinde tutarlılık oranlarını (CR) elde edilerek yapılan karşılaştırmaların doğru ve sağlıklı sonuçlar vermesini arzu ederiz. Tutarlılık oranları bize bu açıdan bilgi vermektedir.

Tutarlılık oranının elde edilmesi için aşağıda verilen adımlar izlenir:

- M karşılaştırma matrisi ile V öncelik vektörü çarpılmıştır ve elde edilen yeni sütun vektörü Z ile gösterilmiştir.
- Z sütun vektörü ile V sütun vektörü değerleri karşılıklı bölünerek temel değer (E) elde edilmiştir. Temel değerlerin elde edilmesi için formül (4.4)'ten yararlanılmıştır.
- Temel değerlerin aritmetik ortalaması alınarak öz değer yani (λ) elde edilmiştir. Öz değer hesaplanması için formül (4.5)'ten yararlanılmıştır.
- Öz değerler hesaplandıktan sonra tutarlılık göstergesi (CI) değeri formül (4.6) kullanılarak bulunmuştur.
- Son aşamada ise tutarlılık göstergesi (CI) ile Tablo 1'de verilen rassal gösterge (RI) değerlerinden uygun olan değer seçilerek birbirlerine bölünmüş ve tutarlılık oranı elde edilmiştir. Tutarlılık oranının hesaplanması için formül (4.7)'den faydalanılmıştır. Rassal göstergede uygun değer, kriter sayısına göre seçilir.

$$Z = \begin{bmatrix} m_{11} & m_{12} & \dots & m_{1a} \\ m_{21} & m_{22} & \dots & m_{2a} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ m_{a1} & m_{a2} & \dots & m_{aa} \end{bmatrix} \times \begin{bmatrix} v_1 \\ v_2 \\ \cdot \\ \cdot \\ v_a \end{bmatrix}$$

$$E_i = \frac{m_i}{v_i} \quad (i=1,2,\dots,a) \quad (4.4)$$

$$\lambda = \frac{\sum_{i=1}^a E_i}{a} \quad (4.5)$$

$$CI = \frac{\lambda - a}{a - 1} \quad (4.6)$$

Tablo 1. RI Değerleri (RI Values)

n	1	2	3	4	5	6	7
RI	0	0	0,58	0,9	1,12	1,24	1,32
n	8	9	10	11	12	13	14
RI	1,14	1,45	1,49	1,51	1,48	1,56	1,57

$$CR = \frac{CI}{RI} \quad (4.7)$$

Tutarlılık oranı eğer 0,10'dan küçükse, bu yapılmış olan karşılaştırmaların tutarlı olduğunu göstermektedir. Eğer 0,10'dan büyükse karşılaştırma matrislerinde bir tutarsızlık olduğunu ve yeniden gözden geçirilmesi gerektiği sonucuna varılır.

4.5. Karar Noktalarındaki Sonuç Dağılımının Bulunması (Finding Results in Decision Point Distribution)

Bu bölümde her bir faktör için elde edilmiş V öncelik vektörleri birleştirilerek tek bir matris haline getirilir. Bu matris Q matrisi olarak aşağıda gösterilmiştir.

$$Q = \begin{bmatrix} v_{11} & v_{12} & \dots & v_{1a} \\ v_{21} & v_{22} & \dots & v_{2a} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ v_{b1} & v_{b2} & \dots & v_{ba} \end{bmatrix}$$

Son aşamada ise faktörler için elde edilen öncelik vektörlerinin birleştirilmesi ile oluşturulmuş Q matrisi ve kriterler için elde edilen öncelik vektörünün çarpılması sonucu karar noktalarındaki sonuç dağılımı elde edilmiştir.

Sonuç dağılımı aşağıda verilen T matrisiyle gösterilmiştir.

$$T = \begin{bmatrix} v_{11} & v_{12} & \dots & v_{1a} \\ v_{21} & v_{22} & \dots & v_{2a} \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ v_{b1} & v_{b2} & \dots & v_{ba} \end{bmatrix} \times \begin{bmatrix} v_1 \\ v_2 \\ \cdot \\ \cdot \\ v_a \end{bmatrix} = \begin{bmatrix} t_{11} \\ t_{21} \\ \cdot \\ \cdot \\ t_{b1} \end{bmatrix}$$

5. LİTERATÜR ARAŞTIRMASI (LITERATURE RESEARCH)

Bu konu ile alakalı literatürde pek çok çalışma vardır. Çalışma alanı geniş olduğundan genel çalışma prensiplerinin dışında çok farklı kısıtlamalar ve modeller kullanarak gerçek hayat problemleri ele alınmıştır. Vardiya çizelgeleme ile alakalı yapılan bazı çalışmalar şunlardır:

Topaloğlu ve Özkarahan [1] çalışmalarında esnek bir planlama kullanılarak bir model geliştirilmiş ve hedef programla ile model çözülmüştür. Aly ve Louly [9] çalışmalarında bir telekomünikasyon şirketinde çalışan mühendislerin vardiyalara atanması, belirli izin günlerinin belirlenmesi ve şirketin çalışma prensipleri doğrultusunda birkaç hedef kısıtını gerçekleştirmek için bir model geliştirmişlerdir. Azaiez ve Al Sharif [10] çalışmalarında 0-1 liner programlama modeli kullanılarak hemşire çizelgeleme problemi geliştirmişlerdir. Geliştirilen modelde hem hastanenin istekleri hem de hemşirelerin tercihleri dikkate alınmıştır. Bektur ve Hasgül [11] çalışmalarında bir restoranda çalışan personeli kıdem seviyelerine ayırarak, personellere verilecek izinlerde kıdem seviyelerine göre atama yaparak ve tüm personelin isteklerini karşılayacak şekilde haftalık bir planlama gerçekleştirmişlerdir. Problemin çözümünde hedef programlama kullanmışlardır. Chu [12] çalışmasında Hedef Programlama kullanılarak iş gücü planlaması yapmıştır. Çizelgeleme yapılırken hem esnek, tatmin edici bir çalışma koşulları, boşta olan vardiyalar minimize edilmek hedeflenmiştir. Çalışmayı Hong Kong Uluslararası Havayollarındaki personellerin atanması ve planlanması için yapmıştır. Elomri vd. [13] çalışmalarında belirli bir hastanede çalışan stajyer doktorların Onkoloji ve Hematoloji bölümlerine dengeli bir şekilde atanmaları için bir çalışma sunmuşlardır. Hung-Tso vd. [14] çalışmalarında güney Tayvan'da bir mağazada müşteri hizmetleri bölümünde hazırlanan ekip çizelgeleme problemine çözüm bulmayı amaçlamışlardır. Çalışmada müşterilere verilecek hizmet süreleri bulanık küme ile belirlenmiş ve çözüm için hedef programlama kullanmışlardır. Sulak ve Bayhan [15] çalışmalarında bir hastanedeki hemşirelerin 3 vardiya ve farklı çalışma saatlerini dengeli bir şekilde hemşirelerin atanması için bir model belirlemişlerdir. Problemin çözümünde hedef programlama kullanmışlardır. Topaloğlu [16] çalışmasında farklı klinik ortamlarında çalışan stajyer doktorların planlanması için yapmıştır. Stajyer doktorların kıdem seviyeleri dikkate alınarak problem karışık tam sayılı programlama ile çözülmüştür. Azmat vd. [19] çalışmalarında tam sayılı programlama ile vardiya iş gücü çizelgeleme problemi ele alınmıştır. Yıllık çalışma saati İsviçre'nin yasal kanunları kullanılarak uygulanmıştır. Volgenant [20] çalışmasında personellerin kıdem

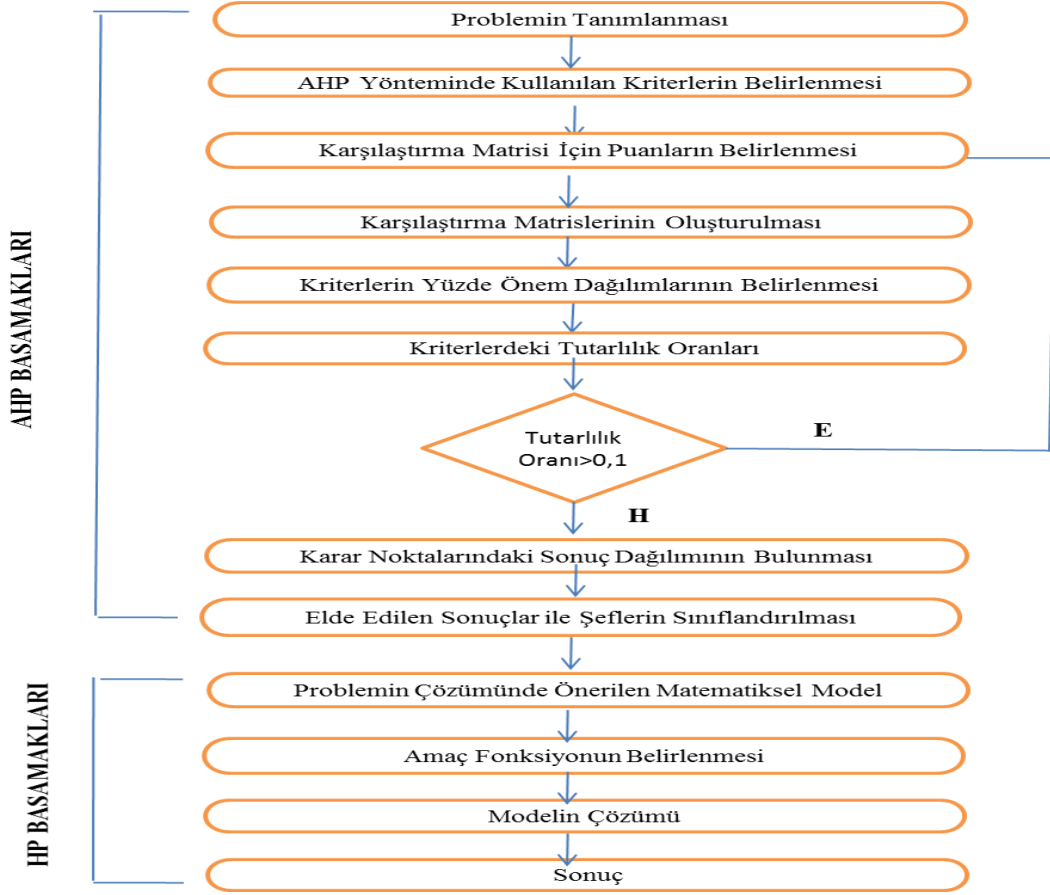
seviyelerine ayırarak nitelikli ve nitelsiz personellerin gerekli iş gücüne atanmasını ele almıştır. Alfares [21] çalışmasında 4 tip izin günleri kısıtları ile üç günlük çalışma haftasında çizelgeleme problemi ortaya koymuştur. Her işçinin verilmiş dört izin gününde en az ikisi ardışık olmalıdır. Hafta sonu tatillerinde yeteri kadar işçi çalıştırmaktır. Maksimum çalışma esnekliği her çalışan için birbirini izleyen dört iş günüdür. Hafta sonu çalışmalarında ardışık haftaların maksimum sayısını belirlemektir. Matematiksel modelde ilk önce problemin yapısı askeri iş gücünü belirlemektir daha sonra iş gücünü kısıtlamak için tam sayılı programlama modeli kullanmıştır. Son olarak birden fazla haftanın çizelgenmesinde hafta sonu çalışmaları ile ardışık hafta sayısını minimize etmek veya sınırlamakta tüm kısıtları karşılamak için bir model oluşturmuştur. Al-Yakoob ve Sherali [22] çalışmalarında Kuveyt Ulusal Petrol şirketinde benzin istasyonlarında çalışan işçilerin atama problemleri üzerinde çalışmışlardır. Kuveyt'teki yaklaşık 86 istasyona işçilerin atanması için çizelgeleme düzenlenmişler ve bu problemi iki aşamalı bir yaklaşım ile benimsemişlerdir, ilk aşama işçilerin istasyonlara atanması ikinci aşama her işçinin tatil günlerinin ve vardiyalarının belirlenmesidir. Bu iki aşamalı yaklaşımda çalışanların tercihleri dikkate alınarak çalışanlar için günlük program sağlanmıştır. Elle yapılan çizelgeleme genellikle düzensizlik, çalışanlar arasında ön yargı ve iş memnuniyetsizliği gibi problemler ortaya çıkmıştır, yeni oluşturulan karar destek modeli ile bu gibi problemlerin önüne geçmeyi amaçlamışlardır. Seckiner vd. [23] çalışmalarında sıkı bir çalışma haftasında hiyerarşik iş gücü sorunu için tam sayılı programlama modeli geliştirmişlerdir. Çalışmada Billionet tarafından geliştirilen tam sayılı programlama kullanarak çözümü elde etmişlerdir. Sinreich ve Jabali [24] çalışmalarında doğrusal optimizasyon ve sezgisel yöntemler simülasyonu tabanlı bir algoritma iş vardiyaları, kaynak planlanması gibi çalışmalar için kullanmışlardır. Sungur [25] çalışmasında her bir çalışma günü ve saati için ihtiyaç duyulan iş gücü sayısını minimum maliyetle karşılamak şartıyla, her bir tura atanarak iş gücü sayısının belirlenmesini amaçlamıştır. Çalışma bir güzellik salonunda ele alınmış ve problemin çözümünde karma tam sayılı programlama kullanmıştır. Sungur [26] çalışmasında Billionet Modeli kullanılarak hiyerarşik iş gücü çizelgeleme problemi çözmek istemiştir. Rong [27] çalışmasında birçok literatürde ve çalışmada olan haftalık iş günlerinin planlanmasının aksine, hafta sonu tatil günlerini dikkate alarak bir tur çizelgeleme oluşturmak istemiştir. Problemin çözümünde tam sayılı programlama kullanmıştır. Bağ vd. [28] çalışmalarında hemşire çizelgeleme problemini ele almışlardır. Problemin çözümünde tamsayı hedef programlama kullanmışlar ve hedeflerin ağırlıklarını belirlemek için analitik ağ prosesi (ANP) kullanmışlardır. Çalışmayı Kırıkkale'deki bir devlet hastanesinde uygulamışlardır. Kassa ve Tizazu [29] çalışmalarında Avanti Blue Nil otelinde çalışan 5 personel için haftalık çalışma günlerini belirlemek adına tam sayılı programlama kullanmışlardır. Öztürkoğlu ve Çalışkan [30] çalışmalarında hemşirelere çalışma saatlerinin başlangıcında esneklik sağlayarak hemşirelerin

memnuniyetini artırmak ve aylık işgücü planının yapılması için bir tam sayılı programlama modeli önermişlerdir. Hidri ve Labidi [31] çalışmalarında yoğun bakım ünitesindeki 3 blokta çalışan hemşireleri belirli gruplara ayırarak ihtiyaç duyulan hemşire sayılarını atama problemini çözmek adına farklı kısıtları kullanarak bir tam sayılı programlama modeli geliştirmişlerdir. Unal ve :

Eren [32] çalışmalarında hedef programlama kullanarak nöbet çizelgeleme problemini ele almışlardır.

6. UYGULAMA (APPLICATION)

Bu çalışmada kullanılan akış şeması aşağıda verilmiştir



Şekil 1. AHP ve hedef programla yöntemiyle vardiya atama problemi
(Shift assignment problem with AHP and goal programming methods)

6.1. Problemin Tanımlanması (Defining the Problem)

Bu çalışmada, 3 vardiya üretim yapan bir fabrikada çalışan şefleri ilk olarak Analitik Hiyerarşi Prosesi (AHP) yönteminden faydalanarak belirli kriterler yardımıyla kıdem seviyelerine ayırdıktan sonra fabrikanın her bir vardiya için ihtiyaç duyduğu işgücünü karşılamak, planlanan dönem boyunca şeflere verilen izin günlerini düzenlemek, her bir şefin toplam işgücünü dengeli ve adaletli bir şekilde dağıtmak ve en önemlisi kıdem seviyelerini kullanarak niteliği çok olan şefler ile niteliği az olan şefleri aynı vardiyaya olabildiğince atamak gibi problemlerin çözümü için bir model geliştirilmeye çalışılmıştır.

Fabrikanın günlük çalışma periyodunda 3 vardiya vardır. Bunlar Sabah-S, Akşam-A, Gece-G olarak

adlandırılmıştır. Her bir vardiyanın saat dilimleri ise Sabah (08:00-16:00), Akşam (16:00-00:00), Gece (00:00-08:00). Sabah vardiyası için 3 şef, Akşam vardiyası için 3 şef, Gece vardiyası için de 1 şef atanmalıdır ve geriye kalan şefler ise ilgili günde izinlidir. Çalışmada Kasım ayındaki personel atamaları gerçekleştirilmiştir.

6.2. AHP Yönteminde Kullanılan Kriterlerin Belirlenmesi (Determination of the criteria used in AHP Method)

AHP analizinde kullanılan kriterler, fabrikada çalışan en az 10 yıl tecrübeli 3 atölye şefinin görüşleri dikkate alınarak belirlenmiş ve değerlendirilmiştir. Çalışmada şefleri becerilerine göre sınıflandırmak için kullanılan kriterler ve şeflerin bu kriterler için kullanılan bilgiler Tablo 2' de verilmiştir.

Tablo 2. Kriterler ve Şeflere Ait Olan Bilgiler
(It belongs to the criteria and Chief Information)

Şef / Kriterler	Tecrübe Yılı	Sertifika ve Eğitim	Aile Yapısı	Fabrikadaki Çalıştığı Yıl	İletişim Becerisi
1	10 yıl	2 adet	2 çocuk	3 yıl	5 puan
2	10 yıl	2 adet	0 çocuk	2 yıl	3 puan
3	14 yıl	3 adet	2 çocuk	6 yıl	7 puan
4	7 yıl	2 adet	4 çocuk	4 yıl	1 puan
5	16 yıl	4 adet	1 çocuk	14 yıl	5 puan
6	9 yıl	1 adet	0 çocuk	8 yıl	9 puan
7	17 yıl	3 adet	2 çocuk	11 yıl	5 puan
8	7 yıl	2 adet	3 çocuk	3 yıl	3 puan
9	7 yıl	1 adet	1 çocuk	1 yıl	1 puan

Tecrübe Yılı: Her şefin toplam çalıştığı yılı göstermektedir.

Sertifika ve Eğitim: Her şef için iş hayatlarında aldıkları sertifikalar ve katıldıkları eğitim kurslarının toplamıdır.

Aile Yapısı: Şeflerin sahip olduğu çocuk sayısını göstermektedir.

Fabrikadaki Çalıştığı Yıl: Şeflerin sadece ilgili fabrikada çalıştığı yılı göstermektedir.

İletişim Becerisi: Şeflerin diğer çalışan personeller ile anlaşılabilir ve çevresindekilere olan davranışları gibi konulardaki becerilerini puanlama yaparak dikkate alınmasıdır. Bu puanlama yapılırken şeflerden sorumlu olunan üst yönetimden bilgi alınarak değerlendirilmiştir.

6.3. Karşılaştırma Matrisi İçin Puanların Belirlenmesi
(Determination of points for the Comparison Matrix)

Aşağıda verilen Tablo 3'te, Tablo 2' de verilen şeflere ait olan bilgiler karşılaştırma matrislerinde kullanılması için puanlara dönüştürülmüştür. Bu tabloda sadece iletişim beceresi kriterinde alınan değerler aynen kullanılmıştır.

Tablo 3. Şeflerin Kriterlere Göre Aldıkları Puanlar (Points they acquired according to the criteria of the Chief)

Tecrübe Yılı	Puan	Sertifika ve Eğitim	Puan	Aile Yapısı	Puan	Fabrikada Çalıştığı Yıl	Puan	İletişim Becerisi	Puan
6 'dan az	1	0 ile 1 arası	3	0 ile 1 arası	3	0 ile 2 arası	1		1
7 ile 9 arası	3	2 ile 3 arası	5	2 ile 3 arası	5	3 ile 5 arası	3		3
10 ile 11 arası	5	4 ile 5 arası	7	4 ile 5 arası	7	6 ile 8 arası	5		5
12 ile 14 arası	7	5 'ten fazla	9	5 'ten fazla	9	9 ile 11 arası	7		7
14 'ten fazla	9					12' den daha fazla	9		9

6.4. Karşılaştırma Matrislerinin Oluşturulması

Şefler için kullanılan kriterlerin kendi aralarında ve her bir kriter için bütün şefler dikkate alınarak EK1'de karşılaştırılmaları gösterilmiştir. Bu karşılaştırmalar yapılırken Saaty' nin 1990'da kullandığı önem skalası ile kullanılmıştır.

Karşılaştırma matrislerinde Tablo 2'deki bilgiler, Tablo 3'teki puanlamalara dönüştürülerek oluşturulmuştur.

6.5. Kriterlerdeki Tutarlılık Oranları

Tablo 4'te AHP yönteminde kullanılan kriterin Tutarlılık oranları verilmiştir.

Tablo 4. Tutarlılık Oranları (Consistency Ratios)

	Tutarlılık Oranları
Kriterler Arasında	0,05
Tecrübe Yılı	0,02
Sertifika ve Eğitim Kursları	0,02
Aile Yapısı	0,04
Fabrikadaki Çalıştığı Yıl	0,03
İletişim Becerisi	0,03

6.6. Karar Noktalarındaki Sonuç Dağılımının Bulunması (Finding Results in Decision Point Distribution)

Bu aşama karar noktalarında oluşturulan öz vektörler birleştirilerek ve kriterler için elde edilen öz vektör çarpılarak sonuç dağılımı elde edilir. Çalışmada elde edilen sonuçlar aşağıda verilmiştir. Tablo 4'teki sonuçlara göre her bir kriter için elde edilen tutarlılık oranı 0,10'dan küçüktür. Dolayısıyla tutarlı bir sonuç elde edilmiştir.

Tablo 5. Şeflerin Becerilerine Göre Aldıkları Puanlar
(Points they received by Chief of skills)

Şefler	Sonuç	Aldıkları Puan
1	0,09	3
2	0,07	3
3	0,14	5
4	0,06	2
5	0,27	10
6	0,07	2
7	0,21	8
8	0,06	2
9	0,03	1

Tablo 5'te puanlar belirlenirken 1 ile 10 arası puanlama yapılması için her bir değer ilk olarak 100 ile daha sonra da 10/27 ile çarpılmıştır.

6.5. Elde Edilen Sonuçlar ile Şeflerin Sınıflandırılması (Classification of the Chief Obtained Results)

Tablo 5'ten yararlanılarak şefler şu şekilde sınıflandırılmıştır.

Tablo 6. Niteliklerine Göre Şeflerin Sınıflandırılması
(Classification of Chief according to their qualities)

Nitelikler	Şefler	Puan
Çok Nitelikli	5, 7, 3	5
Orta Nitelikli	1, 2	3
Az Nitelikli	4, 6, 8, 9	1

Oluşturulan matematiksel modelde Tablo 6'teki puanlar kullanılmaktadır. Yeniden bir puan verilmesinin sebebi modelin çözümünde istenilen, niteliği çok olan şefler ile az olan şefler aynı vardiyaya atanması probleminin daha sağlıklı ve istenilene uygun olabilmesinden dolayı yeniden bir puanlama gereği duyulmuştur.

6.7. Problemin Çözümünde Önerilen Matematiksel Model (Suggested Mathematical Model of Problem Solution)

Parametreler:

n : Fabrikadaki şef sayısı	$n=9$
m : Gün sayısı	$m=30$
l : Vardiya sayısı	$l=3$
i : Şef indeksi,	$i=1,2,\dots,n$
j : Gün indeksi,	$j=1,2,\dots,m$
k : Vardiya indeksi	$k=1,2,\dots,l$
a_i : Şeflerin aldığı puanlar	$i=1,2,\dots,n$

Karar Değişkenleri:

$$X_{ijk} = \begin{cases} 1, & i. \text{şef } j. \text{gündeki } k. \text{ vardiyaya atanırsa} \\ 0, & \text{diğer durumlarda} \end{cases}$$

$$h_{ij} = \begin{cases} 1, & i. \text{şef } j. \text{günde izinli ise} \\ 0, & \text{diğer durumlarda} \end{cases}$$

Kısıtlar:

1.Kısıt: Günlük personel ihtiyacını karşılama kısıtları:

Sabah vardiyası için ihtiyaç duyulan şef sayısı

$$\sum_{i=1}^n X_{ij1} = 3 \quad j=1,2,\dots,m$$

Akşam vardiyası için ihtiyaç duyulan şef sayısı

$$\sum_{i=1}^n X_{ij2} = 3 \quad j=1,2,\dots,m$$

Gece vardiyası için ihtiyaç duyulan şef sayısı

$$\sum_{i=1}^n X_{ij3} = 1 \quad j=1,2,\dots,m$$

2.Kısıt: Her şefi günde sadece bir vardiya atama kısıtı:

$$\sum_{k=1}^l X_{ijk} \leq 1 \quad i=1,2,\dots,n \quad j=1,2,\dots,m$$

3.Kısıt: Her şefin izinli olduğu gün çalışmaması kısıtı:

$$\sum_{k=1}^l X_{ijk} \leq (1 - h_{ij}) \quad i=1,2,\dots,n \quad j=1,2,\dots,m$$

4.Kısıt: Her şef ardı ardına 5 günden fazla çalışmaması kısıtı:

$$h_{ij} + h_{i(j+1)} + h_{i(j+2)} + h_{i(j+3)} + h_{i(j+4)} + h_{i(j+5)} \geq 1 \quad i=1,2,\dots,n \quad j=1,2,\dots,m-5$$

5.Kısıt: Her şefin planlama dönemi boyunca en fazla çalışması gereken vardiya sayılarının belirlenmesi kısıtları:

$$\sum_{j=1}^m X_{ij1} \leq 10 \quad i=1,2,\dots,n$$

$$\sum_{j=1}^m X_{ij2} \leq 10 \quad i=1,2,\dots,n$$

$$\sum_{j=1}^m X_{ij3} \leq 4 \quad i=1,2,\dots,n$$

6.Kısıt: Her şefin planlama dönemi boyunca en az çalışması gereken vardiya sayılarının belirlenmesi kısıtları:

$$\sum_{j=1}^m X_{ij1} \geq 9 \quad i=1,2,\dots,n$$

$$\sum_{j=1}^m X_{ij2} \geq 9 \quad i=1,2,\dots,n$$

$$\sum_{j=1}^m X_{ij3} \geq 3 \quad i=1,2,\dots,n$$

7.Kısıt: Herhangi bir şef herhangi bir günde gece vardiyasına atandığı takdirde ertesi gün ki sabah ve akşam vardiyalarına atanmaması kısıtı:

$$X_{ij3} + X_{i(j+1)1} + X_{i(j+1)2} \leq 1 \quad i=1,2,\dots,n \quad j=1,2,\dots,m$$

8.Kısıt: Herhangi bir şef herhangi bir günde akşam vardiyasına atandığı takdirde ertesi gün ki sabah vardiyasına atanmaması kısıtı:

$$X_{ij2} + X_{i(j+1)1} \leq 1 \quad i=1,2,\dots,n \quad j=1,2,\dots,m$$

Hedef Kısıtları:

Modelde belirtilen kısıtlara gevşek değişkenler eklenerek istenilen hedeflerin gerçekleşmesi ve sapmaların en küçüklenmesi amaçlanmıştır Modelde 5 tane hedef kısıtı vardır.

Hedef 1: Şefler vardiyalara atanırken izin günü-çalışma günü-izin günü olarak atanması en aza indirgenmesi istenen hedef kısıtı:

$$h_{ij} + X_{i(j+1)1} + X_{i(j+1)2} + X_{i(j+1)3} + h_{i(j+2)} + d1_{ij}^- - d1_{ij}^+ = 2 \quad i=1,2,\dots,n \quad j=1,2,\dots,m-2$$

Hedef 2: Şefler vardiyalara atanırken çalışma günü-izin günü-çalışma günü olarak atanması en aza indirgenmesi istenen hedef kısıtı:

$$X_{ij1} + X_{ij2} + X_{ij3} + h_{i(j+1)} + X_{i(j+2)1} + X_{i(j+2)2} + X_{i(j+2)3} + d2_{ij}^- - d2_{ij}^+ = 2 \quad i=1,2,\dots,n \quad j=1,2,\dots,m-2$$

Hedef 3: Planlama dönemi boyunca her şefin atandığı toplam vardiya sayıları mümkün olduğunca eşit olması istenen hedef kısıtı:

$$\sum_{j=1}^m (X_{ij1} + X_{ij2} + X_{ij3}) + d3_{ij}^- - d3_{ij}^+ = 23 \quad i=1,2,\dots,m$$

Hedef 4: Planlama dönemi boyunca günlük sabah vardiyasına atanan şefler arasında niteliği çok olan ile niteliği az olan şefler mümkün olduğunca aynı vardiyaya atanmasını hedefleyen kısıt:

$$\sum_{i=1}^m (a_i * X_{ij1}) + d4_{ij}^- - d4_{ij}^+ = 9 \quad j=1,2,\dots,m$$

Hedef 5: Planlama dönemi boyunca günlük akşam vardiyasına atanan şefler arasında niteliği çok olan ile niteliği az olan şefler mümkün olduğunca aynı vardiyaya atanmasını hedefleyen kısıt:

$$\sum_{i=1}^m (a_i * X_{ij2}) + d5_{ij}^- - d5_{ij}^+ = 9 \quad j=1,2,\dots,m$$

Amaç Fonksiyonu:

$$MinZ = \sum_{i=1}^n \sum_{j=1}^m (d3_{ij}^- - d3_{ij}^+) + (d2_{ij}^- - d2_{ij}^+) + (d1_{ij}^- - d1_{ij}^+) + \sum_{j=1}^m d4_j^- + d5_j^-$$

Modelin çözümünde "Intel (R) Core (TM) i5-3210 M CPU@2.50 GH" işlemcisi, 8 GB belleği ve Windows 10 işletim sistemine sahip bilgisayar kullanılmıştır. İlgili verilerin girilmesiyle model ILOG CPLEX Studio IDE programında yazılmış ve CPLEX çözücüsü ile çözülmüştür.

Şef/ Gün	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	S	A	G	Toplam	
1	S	S		S	A	G		S	A	A	G	G		S	G		S	A	A		S	S	S	S	A		A	A	A	A	10	10	4	24	
2		A	A		S	S	S	G		S	S		S	A	A	A		S	S	A	A	A		S	A	G	G		S	10	10	3	23		
3	A	A		S	A	A		S	S	S	S	A		S	S	G		S	S	A	A	A		G	G	G		S	A	A	10	10	4	24	
4	A	G	G	G		S	S	S	S	A		S	A	A	A		S	A	A			A	A		S	S	S	S		10	10	3	23		
5	S	S	S		S	A	A			A	A		S	A	A		A	A	G	G	G		S	S	S	S	S		A	A	10	10	3	23	
6	G		S	A	A	A	G		S	S	A	A	A		S	S	S	S	A		S	A	G		A	A		S	S		10	10	3	23	
7	A		A	A		S	S	A	G	G		S	A	G		S	S	A		S	S	S	A	A		A	A		S	S	10	10	3	23	
8		S	S	S	S		A	A	A		A	A	G		S	S	A	G		A	A	G		S	S	S	A	A		S	10	10	3	23	
9	S	A	A	A	G		A	A	A		S	S	S	S		A	G		S	S		S	S	A	A		S	A	G	G	10	10	4	24	
S	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	
A	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
G	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

Şekil 2. Şefler için oluşturulan çalışma çizelgesi (The work schedule is created for Chief)

Vardiyalar S:Sabah A:Akşam G:Gece

Sabah	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	
Ç.N	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	
O.N	1	1		1	1	1	1	1		1	1		1	1		1		1	1	1	1	1	1	1	1				1		
A.N	1	1	2	1	1	1	1	1	2	1	1	2	1	1	2	2	1	2	1	1	1	1	1	1	1	1	2	2	2	2	1

Akşam	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
Ç.N	2	1	1	1	1	2	1	1		1	1	1	1	1	1		1	2		1	1	1	1	1		1	1		2	2
O.N		1	1		1				1	1				1	1	1	1	1	1		1	1	1		1	1	1	1	1	1
A.N	1	1	1	2	1	1	2	2	2	1	2	2	2	1	1	2	1		2	2	1	1	1	2	2	1	1	2		

Şekil 3. Niteliklerine göre sınıflandırılan şeflerin günlük atandığı vardiyalar (The chef is assigned daily shifts are classified according to their quality)

Nitelikler Ç.N: Çok Nitelikli **O.N:** Orta Nitelikli **A.N:** Az Nitelikli

Vardiya	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
S	9	9	7	9	9	9	9	9	7	9	9	7	9	9	7	7	9	7	9	9	9	9	9	9	9	7	7	7	7	9
A	11	9	9	7	9	11	7	7	5	9	7	7	7	9	9	5	9	13	5	7	9	9	9	7	5	9	9	5	13	13

Şekil 4. Niteliklerine göre sınıflandırılan şeflerin günlük atandığı vardiyalardaki toplam puanları (Total points in the assigned day shift supervisor of classified according to their quality)

Vardiyalar S: Sabah **A:** Akşam

Yapılan çalışmanın sonuçları Şekil 2, Şekil 3 ve Şekil 4'te verilmiştir.

Şekil 2'de gösterilen pembe renkteki şefler çok nitelikli, yeşil renkteki şefler orta nitelikli, sarı renkte gösterilen şefler ise az nitelikli şefleri göstermektedir. Bu çizelgede her gün ihtiyaç duyulan şef sayıları, her şefin aylık atandığı vardiyalar ve toplam vardiya sayıları gösterilmiştir. Şeflerin dengeli ve adaletli bir şekilde atanması hedefine ulaşılmıştır.

Şekil 3'te AHP yöntemi ile niteliklerine göre sınıflandırılan şeflerin sabah ve akşam vardiyalarına niteliği çok olan ile niteliği az olan şefler aynı vardiyaya atanması hedefinin olabildiğince sağlandığını göstermektedir.

Şekil 4'te sabah ve akşam vardiyasına atanan şeflerin puanlarının toplamıdır yani çok nitelikli şef 5 puan, orta nitelikli şef 3 puan ve az nitelikli şef 1 puan değerlerini baz alarak şeflerin atandığı vardiyaya göre istenilen hedef puanlarının ne kadar gerçekleştiğini göstermektedir.

7. SONUÇLAR (RESULTS)

Bu çalışmada bir fabrikada çalışan şeflerin aylık çalışma süresi boyunca fabrikanın belirlediği vardiyalara dengeli ve adaletli bir şekilde atanması sağlanmıştır ve dahası diğer çalışmalardan farklı olarak AHP yöntemini ile şefleri belirli kriterler kapsamında niteliklerine göre sınıflandırılmıştır. Bu sınıflandırılan şefleri daha sonra günlük sabah ve akşam vardiyalarına niteliklerine göre atamalar yaparak fabrikadaki verimliliğini artırılmasında

ve şeflerin birbirlerini tamamlanması konusunda önemli bir rol oynamıştır.

Çalışmanın çözümü için tam sayılı programlama ve hedef programlama kullanılmıştır ve istenilen 5 hedef kısıtı vardır. Yapılan diğer çalışmalar incelendiğinde önerilen model ile alakalı çok fazla çalışma tespit edilememiştir.

Bundan sonraki çalışmalarda daha farklı kısıtlar kullanılarak fabrikada çalışan başka personellerin çizelgelenmesi yapılabilir. Planlanması yapılan personelin istek ve önerileri modele eklenerek daha farklı çalışmalar yapılabilir.

KAYNAKLAR (REFERENCES)

- [1] S. Topaloglu, I. Ozkarahan, An implicit goal programming model for the tour scheduling problem considering the employee work preferences. *Annals of Operations Research*, 128, 135-158, 2004.
- [2] A. T. Ernst, H. Jiang, M. Krishnamoorthy, B. Owens, D. Sier, An annotated bibliography of personnel scheduling and rostering. *Annals of Operations Research*, 127, 21- 144, 2004.
- [3] G. B. Dantzig, Letter to the editor—A comment on Edie's "Traffic delays at toll booths". *Operations Research*, 2, 339-341, 1954.
- [4] L. C. Edie, Traffic delays at toll booths. *Operations Research*, 2, 107-138, 1954.
- [5] J. Ignizio, Introduction to goal programming, Sage Publications Inc., Beverley Hills, California, USA, 1985.

- [6] A. Charnes, W. W. Cooper, R. Ferguson, Optimal estimation of executive compensation by linear programming, *Management Science*, 1, 138-151, 1955
- [7] A. Charnes, W.W. Cooper, *Management models and industrial applications of linear programming*, Wiley, New York, 1965-66, 1, and 1967, 2, 1961.
- [8] A. Charnes, W.W. Cooper, Goal programming and multiple objective optimizations, *European Journal of Operational Research I*, 39-54, 1977.
- [9] M. Aly, O. Louly, A goal programming model for staff scheduling at a telecommunications center, *J Math Model Algor*, 12,167-178, 2012.
- [10] M. N. Azaiez, S. S. Al Sharif, A 0-1 goal programming model for nurse scheduling. *Computers & Operations Research*, 32, 491-507, 2005.
- [11] G. Bektur, S. Haşgöl, Kıdem seviyelerine göre işgücü çizelgeleme problemi: Hizmet sektöründe bir uygulama. *Afyon Kocatepe Üniversitesi, İİBF Dergisi*,385-402, 2013.
- [12] S. C. K. Chu, Generating, scheduling and rostering of shift crew-duties: Applications at the Hong Kong International Airport. *European Journal of Operational Research*, 177, 1764- 1778, 2007.
- [13] A. Elomri, S. Elthlatiny, Z.S. Mohamed, A goal Programming model for fairly scheduling medicine residents. *Int. J Sup. Chain. Mgt, IJSCM*, 4, 2050-7399, 2015.
- [14] L. Hung-Tso, C. Yen-Ting, C. Tsung-Yu, L. Yi-Chun, Crew rostering with multiple goals: An empirical study. *Computers & Industrial Engineering*, 63, 483-493, 2012.
- [15] H. Sulak, M. Bayhan, A Model suggestion and an application for nurse scheduling problem. *Journal of Research in Business, Economics and Management*, 2395-2210, 2016.
- [16] S. Topaloglu, A shift scheduling model for employees with different seniority levels and an application in healthcare. *European Journal of Operational Research*, 198, 943 957, 2009.
- [17] T. L. Saaty, *The Analytic Hierarchy Process*, McGraw-Hill Inc., New York, 1977.
- [18] T.L. Saaty, How to make a decision: the analytic hierarchy process', *European Journal of Operational Research*, 48, 9-26, 1990.
- [19] C. S. Azmat, T. Hurlimann, M. Widmer, Mixed integer programming to schedule a single-shift workforce under annualized hours. *Annals of Operations Research*, 128, 199-215, 2004.
- [20] A. Volgenant, A note on the assignment problem with seniority and job priority constraints. *European Journal of Operational Research*, 154, 330-335, 2004.
- [21] H. K. Alfares, Compressed workweek scheduling with days-off consecutivity weekendoff frequency and work stretch constraints. *Infor*, 44, 175-189, 2006.
- [22] S. M. Al-Yakoob, H. D. Sherali, Mixed-integer programming models for an employee scheduling problem with multiple shifts and work locations. *Annals of Operations Research*, 155, 119-142, 2007.
- [23] S. U. Seckiner, H. Gokcen, M. Kurt, An integer programming model for hierarchical workforce scheduling problem. *European Journal of Operational Research*, 183, 694-699, 2007.
- [24] D. Sinreich, O. Jabali, Staggered work shifts: a way to downsize and restructure an emergency department workforce yet maintain current operational performance. *Health Care Management Science*, 10, 293-308, 2007.
- [25] B. Sungur, Bir güzellik salonunun tur çizelgeleme problemi için karma tamsayı hedef programlama modelinin geliştirilmesi. *İstanbul Üniversitesi İşletme Fakültesi Dergisi Istanbul University Journal of the School of Business Administration*, 1, 49-64, 2008.
- [26] B. Sungur, Hiyerarşik işgücü çizelgeleme problemi için tamsayı programlama modeli. *Dokuz Eylül Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, 24, 2, 23-31, 2009.
- [27] A. Y. Rong, Monthly tour scheduling models with mixed skills considering weekend off requirements. *Computers & Industrial Engineering*, 59, 334-343, 2010.
- [28] N. Bağ, M. Özdemir, T. Eren, 0-1 hedef programlama ve ANP yöntemi ile hemşire çizelgeleme problemi çözümü. *International Journal of Engineering Research and Development*,1,2-6, 2012.
- [29] B. A. Kassa, A. E. Tizazu, Personnel scheduling using an integer programming model- an application at Avanti Blue-Nile Hotels. *SpringerPlus*, 2:333, 1-7, 2013.
- [30] Y. Öztürkoğlu, F. Çalışkan, Hemşire çizelgenmesinde esnek vardiya planlanması ve hastane uygulaması. *Dokuz Eylül Üniversitesi Sosyal Bilimler Enstitüsü Dergisi* 16,1,115-133, 2014.
- [31] L. Hidri, M. Labidi, Optimal physicians schedule in an Intensive Care Unit. *IOP Conf. Series: Materials Science and Engineering* 131, 1-8, 2016.
- [32] F.M. Unal, T. Eren, Hedef programlama ile nöbet çizelgeleme probleminin çözümü. *Akademik Platform Mühendislik ve Fen Bilimleri Dergisi*, 4, 1, 28-37, 2016.

EK 1:

Kullanılan Kriterler için Karşılaştırma Matrisi (Comparison Matrix for Criteria)

Kriterler	Tecrübe Yılı	Sertifikalar ve Eğitim Kurslar	Aile Yapısı	Fabrikadaki Çalıştığı Yıl	İletişim Becerisi
Tecrübe Yılı	1	3	9	5	7
Sertifikalar ve Eğitim Kursları	0,33	1	7	3	5
Aile Yapısı	0,11	0,14	1	0,2	0,33
Fabrikadaki Çalıştığı Yıl	0,2	0,33	5	1	3
İletişim Becerisi	0,14	0,2	3	0,33	1

Şeflerin Tecrübe Yıllarına Göre Karşılaştırma Matrisi (Comparison Matrix According to Chiefs' Experiences of Years)

Şefler	1	2	3	4	5	6	7	8	9
1	1	1	0,33	3	0,2	3	0,2	3	3
2	1	1	0,33	3	0,2	3	0,2	3	3
3	3	3	1	5	0,33	5	0,33	5	5
4	0,33	0,33	0,2	1	0,14	1	0,2	1	1
5	5	5	3	7	1	7	1	7	7
6	0,33	0,33	0,2	1	0,14	1	0,14	1	1
7	5	5	3	5	1	7	1	7	7
8	0,33	0,33	0,2	1	0,14	1	0,14	1	1
9	0,33	0,33	0,2	1	0,14	1	0,14	1	1

Şeflerin Aldığı Sertifikalara Göre Karşılaştırma Matrisi (Comparison Matrix According to the certificate received the Chiefs)

Şefler	1	2	3	4	5	6	7	8	9
1	1	1	1	1	0,33	3	1	1	3
2	1	1	1	1	0,33	3	1	1	3
3	1	1	1	1	0,33	3	1	1	3
4	1	1	1	1	0,33	3	1	1	3
5	3	3	3	3	1	5	3	3	5
6	0,33	0,33	0,33	0,33	0,2	1	0,33	0,33	1
7	1	1	1	1	0,33	3	1	1	3
8	1	1	1	1	0,33	3	1	1	3
9	0,33	0,33	0,33	0,33	0,2	1	0,33	0,33	1

Şeflerin Aile Yapılarına Göre Karşılaştırma Matrisi (Comparison Matrix of Chiefs by Family Constructs)

Şefler	1	2	3	4	5	6	7	8	9
1	1	3	1	0,33	3	3	1	1	3
2	0,33	1	0,33	0,2	1	1	0,33	0,33	1
3	1	3	1	0,33	3	3	1	1	3
4	3	5	3	1	5	5	3	3	5
5	0,33	1	0,33	0,2	1	1	0,33	0,33	1
6	0,33	1	0,33	0,2	1	1	0,33	0,33	1
7	1	3	1	0,33	3	3	1	1	3
8	1	3	1	0,33	3	3	1	1	3
9	0,33	1	0,33	0,2	1	1	0,33	0,33	1

Şeflerin Fabrikadaki Çalıştığı Yıllara Göre Karşılaştırma Matrisi (Comparison Matrix by years worked as chief of the Factory)

Şefler	1	2	3	4	5	6	7	8	9
1	1	3	0,33	1	0,14	0,33	0,2	1	3
2	0,33	1	0,2	0,33	0,11	0,2	0,14	0,33	1
3	3	5	1	3	0,2	1	0,33	3	5
4	1	3	0,33	1	0,14	0,33	0,2	1	3
5	7	9	5	7	1	5	3	7	9
6	3	5	1	3	0,2	1	0,33	3	5
7	5	7	3	5	0,33	3	1	5	7
8	1	3	0,33	1	0,14	0,33	0,2	1	3
9	0,33	1	0,2	0,33	0,11	0,2	0,14	0,33	1

Şeflerin İletişim Becerilerine Göre Karşılaştırma Matrisi (Comparisons of Chiefs' Communication Skills)

Şefler	1	2	3	4	5	6	7	8	9
1	1	3	0,33	5	1	0,2	1	3	5
2	0,33	1	0,2	3	0,33	0,14	0,33	1	3
3	3	5	1	7	3	0,33	3	5	7
4	0,2	0,33	0,14	1	0,2	0,11	0,2	0,33	1
5	1	3	0,33	5	1	0,2	1	3	5
6	5	7	3	9	5	1	5	7	9
7	1	3	0,33	5	1	0,2	1	3	5
8	0,33	1	0,2	3	0,33	0,14	0,33	1	3
9	0,2	0,33	0,14	1	0,2	0,11	0,2	0,33	1

İlişkisel ve İlişkisel Olmayan (NoSQL) Veri Tabanı Sistemleri Mimari Performansının Yönetim Bilişim Sistemleri Kapsamında İncelenmesi

Serdar ÖZTÜRK¹, Hatice Ediz ATMACA²

¹Yönetim Bilişim Sistemleri, Bilişim Enstitüsü, Gazi Üniversitesi, Ankara, Türkiye

²Endüstri Mühendisliği, Mühendislik Fakültesi, Gazi Üniversitesi, Ankara, Türkiye

serdarozturktr@hotmail.com, hediz@gazi.edu.tr

(Geliş/Received: 14.10.2016; Kabul/Accepted: 25.05.2017)

DOI: 10.17671/gazibtd.309303

Özet— Veri tabanları, veri fazlalığını kontrol eden ve veri tutarlılığını koruyan en önemli sistemlerden biridir. Veri entegrasyonu ile verilere basit şekilde ulaşılması, düzenlenmesi ve paylaşılması gibi avantajlar sağlamaktadır. Veri depoları oluşturulması sırasında yaygın olan yaklaşım, verilerin uygulama modeline yakın bir model ile saklanması üzerine kurgulanmaktadır. Bu çalışmada öncelikle ilişkisel ve ilişkisel olmayan veri tabanlarının yönetim bilişim sistemleri kapsamında öğrenilmesi ve niteliklerinin belirlenmesi ile ilgili bilgiler sunulmuştur. Sonrasında veri tabanı modelinin seçilmesi ve ölçütlerinin ortaya koyulmasıyla en uygun performans ölçümleri, veri tabanlarının avantaj ve dezavantajları ile oluşturulacak yapının uygun hale getirilmesi hakkında bilgiler verilmiştir. Daha sonra ilişkisel veri tabanı ve dağıtık veri tabanlarının performans karşılaştırması yer almaktadır. Çalışmanın son bölümünde ise veri tabanları yapıları karşılaştırmalı olarak incelenerek sonuç ve öneriler sunulmuştur.

Anahtar Kelimeler— İlişkisel veri tabanı, dağıtık veri tabanı sistemleri, ilişkisel olmayan veri tabanı, NoSQL

The Examination of Relational and Non-relational (NoSQL) Database System's Architectural Performances in Terms of Management of Information Systems

Abstract— Databases are one of the most important systems that control the exceed data and protect the data consistency. They provide advantages such as accessing data easily by data integration, arranging data and sharing them. The prevalent approach while creating databases is based on preserving data with a model that is similar to that of applying data. In this work, information about the learning of relational and non-relational database in the perspective of management information systems and identifying the attributes of them is presented primarily. Then, information about the selection of the database model and identifying the criteria, the most suitable performance measures, advantages and disadvantages of the databases and compiling the created structure is expressed. Subsequently, comparison of the performance of the relational and non-relational database systems is explained. Lastly, structures of databases are examined by comparing each other and results and suggestions are presented.

Keywords— Relational database, distributed database systems, non-relational database system, NoSQL

1. GİRİŞ (INTRODUCTION)

Bilgisayar ve iletişim teknolojilerinde yaşanan hızlı gelişim her geçen gün daha fazla organizasyonu etkileyerek farklı çözümler üretmeye zorlamaktadır. Belli başlı bir amaca ulaşmak için veri veya ham bilginin işlenerek ilgililere yarar sağlayacak biçime dönüştürülmüş hali olan bilgi, organizasyonlar tarafından sürekli daha kısa sürede erişilmek istenen en etkili faktör haline

gelmiştir. Bilgisayarlar karar alma sürecinde etkin olarak kullanılarak, “bilgi sistemleri” günümüz trend konuları arasında yerini almıştır [1].

Günümüzde yaşanan bu değişim ve gelişim, verilerin modellenerek saklanması ve dolayısıyla veri tabanı kullanımını zorunlu kılmaktadır. Temel bir kurum rehberinden, orta ve büyük ölçekli işletmelerin kurumsal ve ticari bilgilerinin organize edilerek saklanmasına kadar farklı alanlarda veri modelleme ve depolama gerekliliği

ortaya çıkmaktadır. Verinin büyüklüğü, miktarı ve karmaşıklığı gibi etkenlere bağlı olarak farklı veri modelleme, veri depolama ve sorgulama yöntemleri geliştirilmiştir.

Bu kapsamda, okuma ve yazma gibi işlemlerin yoğun olarak kullanıldığı veri tabanlarında ilişkisel veri tabanlarının yanı sıra ilişkisel olmayan veri tabanı yönetim sistemleri de kullanılmaktadır. Performans ve esneklik özellikleri ile ilişkisel olmayan veri tabanı yönetim sistemleri (NoSQL) eBay ve Amazon gibi dünyaca ünlü şirketler tarafından tercih edilebilir hale gelmiştir [2].

Bu çalışmada “bilişim sistemleri” ve “veri tabanı” kavramları incelenerek ilişkisel ve ilişkisel olmayan veri tabanı yönetim sistemleri mimari performansının detaylı karşılaştırılması yapılmıştır.

2. BİLİŞİM SİSTEMLERİ VE YÖNETİMİ (INFORMATION SYSTEMS AND MANAGEMENT)

Bilişim sistemi, organizasyonlarda karar verme aşamasına kadar bilgiyi toplamak, düzenlemek, işlemek ve saklamak olarak tanımlanabilir. Bilişim sistemlerinde üç aktivite bilgiyi üretmek için gereklidir. Bu aktiviteler: girdi, işlem ve çıktıdır. Girdi, organizasyonun içinden veya dış çevresinden, ham bilgileri (veriyi) toplamaktır. İşlem, bu ham veriyi daha anlamlı biçime çevirir. Çıktı, işlenmiş bilgiyi (enformasyon), insanlara veya kullanılacak olan aktivitelere aktarır.

İşletmeler açısından bilişim sistemleri, herhangi bir girdiyi işlemlere tabi tutup çıktı sağlayan mekanik yapılardan daha fazla anlam ifade etmektedir. Bilişim sistemleri, bilişim teknolojileri altyapısından yararlanan yönetsel çözümlerdir. Bilişim sistemlerini etkin bir şekilde kullanmak için organizasyon, yönetim ve teknolojiye hâkim olmak gerekmektedir [3].



Şekil 2.1 Bilişim Sistemleri Bileşenleri
(Information Systems Components)

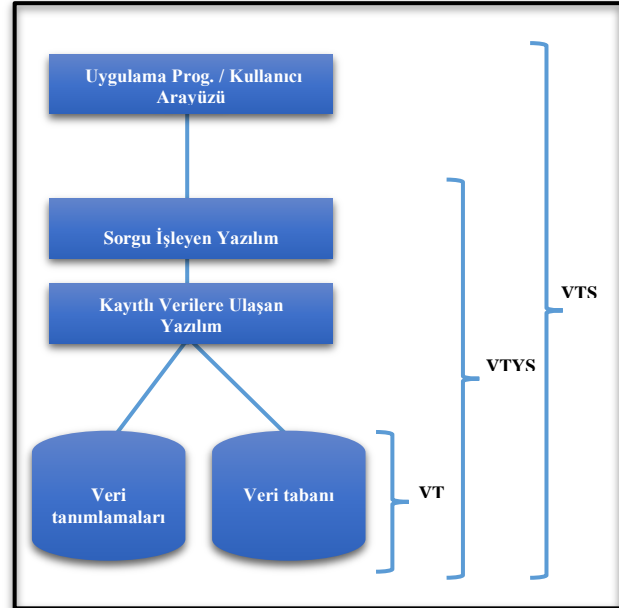
3. VERİ TABANI VE VERİ TABANI YÖNETİM SİSTEMLERİ (DATABASE AND DATABASE MANAGEMENT SYSTEM)

Veri tabanı en genel tanımıyla, kullanım amacına uygun olarak düzenlenmiş veriler topluluğudur. Birbirleriyle ilişkileri olan verilerin tutulduğu, mantıksal ve fiziksel olarak tanımlarının olduğu bilgi depolarıdır. Veri tabanları gerçekte var olan ve birbirleriyle ilişkisi olan nesnelere ve ilişkileri modeller [4].

Veri tabanı yönetim sistemleri (VTYS), verilere aynı anda birden çok bağlantı sağlayabilme özelliği sağlar. Bu sistemler, veri tabanı yönetiminin bir parçası olarak, verinin nasıl depolanacağı, kullanılacağı ve erişileceğini mantıksal olarak yönlendiren bir kurallar sistemidir.

Veri tabanı, VTYS ve uygulama programlarını ile kullanıcı ara yüzlerini içeren yapıya “veri tabanı sistemi (VTS)” denir. Veri tabanı, veri tabanı yönetim sistemi ve veri tabanı sistemi arasındaki ilişki ve işlevler Şekil 3.1’de gösterilmiştir.

Şekil 3.1 VT-VTYS-VTS Arasındaki İlişki ve İşlevler
(Relation and Functions between DB-DBMS-DBS)



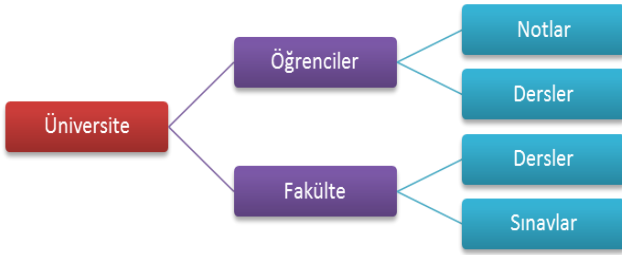
Veri tabanı modellerini sekiz kategoriye ayırabiliriz:

- Düz model veya tablo modeli: İki boyutlu veri grubundan oluşur. Sütunlarda verilerin benzer özellikleri, satırlarda ise veri grupları yer alır. Kullanıcı adlarının ve şifrelerinin tutulduğu veri tabanı buna örnek olarak verilebilir. Böyle bir veri tabanında her satırda bir kullanıcıya ait şifre bilgileri, sütunlarda ise tipleri aynı olan veriler yer alır. Düz veri modeli tek tablodan oluşan bir model olarak düşünülebilir [5].

	Ad Soyad	Kullanıcı Adı	Parola
Kayıt 1	Murat ERGİN	Mergin	kjVdb125
Kayıt 2	Ayşe YILMAZ	Ayılmaz	Bks46db7
Kayıt 3	Can TÜRK	Cturk	fhG8dbt9

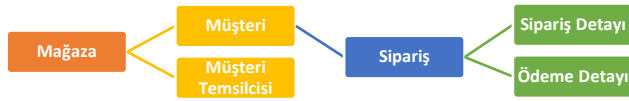
Şekil 3.2 Düz veri modeli örneği
(Instance of flat data model)

- Hiyerarşik Veri Modeli: İlk olarak 1960'lı yıllarda ortaya çıkmış ve adını veriyi depolama yönteminden almıştır. Bu veri tabanının depoladığı yapısal verilere "kayıt" adı verildi. Kayıtlar ağaç mimarisi şeklinde yukarıdan aşağı sıralanmaktadır. Kök adı verilen ilk kaydın bir veya daha çok çocuk kayıtları vardır. Çocuk kayıtlarında kendi çocuk kayıtları olabilir. Kök haricinde bütün kayıtların bir ebeveyni vardır [6].



Şekil 3.3 Hiyerarşik Veri Tabanı Modeli
(Hierarchical Database Model)

- Ağ veri modeli: Bu model 1970'li yılların başında geliştirilmiştir. Hiyerarşik veri modelinin geliştirilmiş halidir. Hızlıca kabul görmesinin nedeni bir verinin doğal olarak başka veriler ile ilişkili olmasıdır. Ağ modelinin hiyerarşik modelden en önemli farkı, uç-düğüm pozisyonundaki verinin iç-düğüme işaret edebilmesidir. Böylelikle ağ modelinde bire-çok ilişkiler yanında, çoka-çok ilişkiler de modellenebilir. Bu veri tekrarını önemli ölçüde azaltır [7].



Şekil 3.4 Ağ Veri Modeli
(Network Data Model)

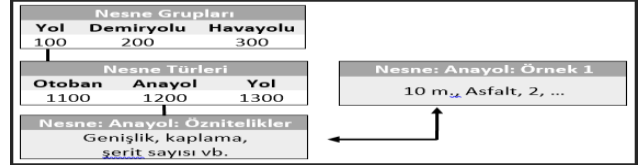
- İlişkisel Veri Modeli: Hiyerarşik ve ağ veri modellerinin, çeşitlenen beklentileri karşılamakta yetersiz kalması, yeni bir model arayışını başlatmış ve ilişkisel veri modeli geliştirilmiştir. E. F. Codd'un 1970'de yazmış olduğu "A Relational Model of Data for Large Shared Data Banks" makalesi ile ilişkisel veri yapılarında büyük bir ilerleme kaydedilmiştir [8-9]. İlişkisel veri modelinin temel kavramı, ilişkidir.

İlişkiler yardımıyla, veri içerisindeki ilişkiler modellenir. Dolayısıyla, ilişkisel bir veri tabanı, çeşitli ilişki örneklerinden oluşur. Kavramsal olarak ilişkiler, satır ve sütunlardan oluşan iki boyutlu tablolarla karakterize edilir. Genellikle veri tabanında her tablo için bir dosya bulunur. Tablonun her satırı birbirisiyle ilişkili verilerin bir topluluğudur. Sütunlarda ise nitelikler bulunur [10].



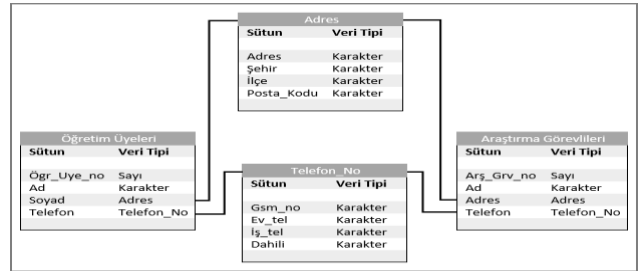
Şekil 3.5 İlişkisel Veri Modeli
(Relational Data Model)

- Nesne Yönelimli Veri Modeli: Daha sonraları ortaya çıkmış ve başarısını kanıtlamıştır. Nesne yönelimli programlamaya dayanan veri modelidir [9].



Şekil 3.6 Nesne Yönelimli Veri Modeli
(Object-Oriented Data Model)

- Nesne İlişkisel Veri Modeli: Nesne ilişkisel veri tabanı, ilişkisel işlevselliğin üzerine nesne yönelimli özellikler içerir. İlişkisel veri tabanları içinde nesne yönelimli karakteristikler içeren ilk veri tabanı 1997 yılında piyasaya sunulan Oracle8'dir.



Şekil 3.7 Nesne İlişkisel Veri Modeli [11]
(Object-Relational Data Model)

- Çoklu Ortam Veri Modeli: Çoklu ortam veri tabanları nesne ilişkisel veri tabanları ile büyük benzerlikler gösterir. Bununla birlikte, film, müzik, metin ve video gibi büyük nesnelere işlemek ve aynı zamanda işleme sırasındaki adımları kullanıcıya göstermemek için farklı özellikler taşır. Çoklu ortam veri tabanlarının desteklemesi gereken üç temel özellik; Veri miktarı, Süreklilik ve Senkronizasyondur. Çoklu ortam veri tabanı uygulaması, imge görüntüleme, uzaktan

görüntülü eğitim, üç boyutlu tıbbi görüntü kayıtları depolanması konularında özellikle tıp bilgi sistemlerinde kullanılmaktadır [6].

- Dağıtık Veri Modeli: Dağıtık veri tabanları, iki ya da daha fazla bilgisayarda depolanan ve bir ağ üzerinde dağıtılan bilgiler için kullanılan veri tabanı grubudur. Veri tabanını ağ üzerinden paralel kullanmak için parçalara ayırmak, sorguların daha hızlı işlenmesini sağlar. Böyle bir sistemde, birden fazla veri tabanına erişilmesine rağmen, kullanıcı bir tek veri tabanı gibi çalışıyormuş gibi işlem yapar [6].

4. VERİ TABANI TASARIMI (DATABASE DESIGN)

Veri tabanı tasarımında; gerçeğin, gereksinim ve beklentiler çerçevesinde modellenerek veri tabanına aktarılması gerekir [10].

Veri tabanı tasarımında ilk olarak, olası veri tabanı kullanıcı gereksinimlerinin belirlenmesi gerekir. Söz konusu gereksinimler, veri tabanında yer alacak veri gruplarını, verilerin tiplerini ve verinin fiziksel olarak depolanması için kullanılacak olan veri yapılarını belirler. Gerçeğin veri tabanındaki sayısal temsili, onun belli bir perspektiften bir modeli olup, bir veri tabanı sisteminde gerek kullanıcılar ve gerekse bilgisayar tarafından anlaşılabilir bir tarzda tanımlanması gerekir. Böyle bir tanımlama, veri tabanı literatüründe “şema” olarak adlandırılır. Kullanıcı ve bilgisayar düzeyleri sırasıyla "kavramsal" ve "fiziksel" düzeyler, bu düzeylerdeki şemalar da “kavramsal şema” ve “iç şema” olarak anılırlar. Kavramsal ve fiziksel düzeylerdeki şemalar, farklı anlayış mekanizmalarına hitap ettiklerinden, kullanılacak veri modelleri de farklı olacaktır. Her iki düzeyde kullanılmak üzere, çeşitli veri modelleri geliştirilmiştir [10].



Şekil 4.1 Veri Tabanı Tasarım Aşamaları
(Database Design Stages)

Geleneksel veri tabanı tasarımı, kullanıcı düzeyinden fiziksel düzeye doğrudur. Kavramsal tasarımda, gereksinimlere göre kavramsal şema belirlenir. Kavramsal şema tanımlamada, kavramsal ya da mantıksal veri modelleri kullanılabilir. Kavramsal şema, ortalama veri tabanı kullanıcısı için, veri tabanının yapısını genel olarak tanımlar. Kullanıcıların veri tabanının yapısını anlamalarına ve böylece uygulamalarını modellemelerini sağlar. Kavramsal şema, fiziksel depolama yapılarının ayrıntılarına girmeden, varlıklar, veri tipleri, varlıklar arasındaki ilişki tipleri ve kısıtlayıcılar üzerinde yoğunlaşır. Bu bakımdan kavramsal şema, yüksek düzeyli bir tanımlamadır. Diğer bir ifadeyle, kavramsal şema, yazılım ve donanımdan bağımsızdır ve son kullanıcı tarafından anlaşılması da daha kolaydır [10].

Kavramsal veri modelleri oldukça yüksek düzeyli olduklarından, kavramsal bir veri modelinde tanımlı bir şema genellikle doğrudan gerçekleştirilemez. Bu nedenle, geleneksel veri tabanı tasarımında, kavramsal tasarımdan sonraki adım, çoğunlukla, gerçekleştirim için kullanılacak bir veri tabanı yönetim sisteminin seçimidir. Öte yandan, bugün piyasadaki veri tabanı yönetim sistemlerinin çoğu mantıksal bir veri modeli kullanmaktadır. O nedenle mantıksal veri modelleri, gerçekleştirim veri modelleri olarak da bilinirler. O halde, kavramsal veri modelindeki kavramsal şema, veri tabanı yönetim sisteminin veri modelinde yeniden tanımlanmalıdır. Buradaki işlem, iki veri modeli arasında bir dönüşüm olup, bazen “mantıksal veri tabanı tasarımı” olarak anılır [10].

Fiziksel tasarım aşamasında, verinin en yüksek verim için, veri tabanında fiziksel olarak nasıl organize edilmesi gerektiği belirlenir. Sonuç, iç şemadır [10].

İç şema depolama yapılarını, kayıt formatlarını, kayıt alanlarını, veri tabanına giriş yol ve yöntemleri ile veri tabanının fiziksel gerçekleştirimini ilgilendiren diğer bütün detayları tanımlar. İç şema tanımlamada, genellikle veri yapıları olarak bilinen, fiziksel veri modelleri kullanılır. İç şema, yazılım ve donanıma bağımlıdır [10].

5. İLİŞKİSEL VE İLİŞKİSEL OLMAYAN (NoSQL) VERİ TABANI SİSTEMLERİ (RELATIONAL AND NON-RELATIONAL DATABASE (NoSQL) SYSTEMS)

5.1 İlişkisel Veri Tabanı (Relational Database System)

Günümüzde en yaygın kullanılan veri tabanı sistemlerinden biridir. Satır ve sütunların meydana getirdiği tablolardan oluşur. Bu tablolar birbiri ile ilişkileri olan tablolardır. Dolayısıyla bir veri tabanında ilişkiden söz edebilmek için en az iki tablonun yer alması ve bu iki tablodaki verilerin birbiri ile bir şekilde ilişkilendiriliyor olması gerekir. Bu şekilde ilişkisel veri tabanları, veri

tabanı denilen büyük dosyalardan oluşur. Her bir tablo, belli yapıya uygun verileri saklamak üzere tasarlanır [12].

ACID; klasik ilişkisel veri tabanı sistemlerinde sağlanan temel özellikler aşağıda sunulmuştur [13]:

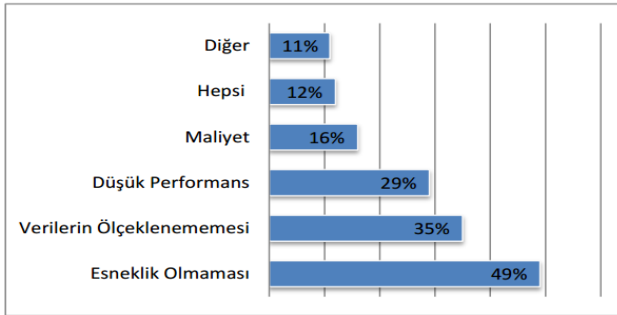
- Bölünmezlik (Atomicity)
- Tutarlılık (Consistency)
- İzolasyon (Isolation)
- Dayanıklılık (Durability)

5.2 İlişkisel Olmayan (NoSQL) Veri tabanı (Non-Relational Database System)

İlişkisel olmayan (NoSQL) veri tabanı; 1998 yılında ilk olarak Carlo Strozzi tarafından öne sürülen bir kavramdır. NoSQL, ilişkisel veri tabanı sistemlerine alternatif bir çözüm olarak ortaya çıkmıştır. İlişkisel olmayan veri tabanları yatay olarak ölçeklendirilen bir veri depolama sistemidir [14].

Dünya'da NoSQL örneklerini incelediğimizde; sosyal ağlarda Digg'in 3 TB'lık çözümü, Facebook'un gelen postaları arama için 50 TB ve eBay'in bütün verileri için 2 PB'lık çözümleri vardır. Veri tabanlarına ilişkin problemlerden biri olan ölçek sorununa, diğer çözümlerin içinde en iyi cevap vereni NoSQL'dir [15]. Günlük 7 TB'lık işlem hacmine sahip Twitter [16] ve 10 TB'lık Facebook örneğindeki gibi, çok büyük verilerin depolanması ve yazılmasında ilişkisel veri tabanlarının eksik kaldığı hususlarda, yatay ölçekleme yapan dağıtık NoSQL çözümleri geliştirilmiştir.

İlişkisel veri tabanı kullanıcılarının, araştırmalar neticesinde NoSQL veri tabanına geçmek istemelerinin nedenleri Şekil 5.1'de yüzde olarak gösterilmiştir.



Şekil 5.1 Neden NoSQL Gerekli
(Why NoSQL Necessary)

Amazon bu gereksinimi "DynamoDB", Google ise "Big Table" ismini verdiği NoSQL veri tabanı sistemi ile çözmektedir. İlişkisel veri tabanını yerine NoSQL veri tabanını tercihi, özellikle hız ve yatay büyüme ile gereksiz ek maliyetten kurtulmaya dayanmaktadır.

İlişkisel veri tabanlarının kullandığı ACID işlemselliğine karşın NoSQL "BASE" (Basically Available- Soft state- Eventually consistent) kısaltması ile ifade edilir.

- Kolay Ulaşılabilirlik (Basically Available): Veri erişim sorunlarını ortadan kaldırmak için kopyaları kullanır ve paylaşılmış ya da bölümlenmiş veriyi birçok sunucudan alır.
- Esnek Durum (Soft state): ACID mantığında veri tutarlılığının olmazsa olmaz bir gereklilik olduğu savunulurdu fakat NoSQL sistemler tutarsız ve süreksiz verilerin barınmasına da izin verir.
- Eninde sonunda Tutarlı (Eventually consistent): Uygulamalar anlık tutarlılıkla ilgili olmasına rağmen, NoSQL sistemlerin gelecekte bir zamanda tutarlı olacağı farz edilir. ACID'in zorunlu tuttuğu tutarlılığa karşın NoSQL'de tanımlanmayan bir zamanda tutarlılığın oluşacağı garanti edilir [17].

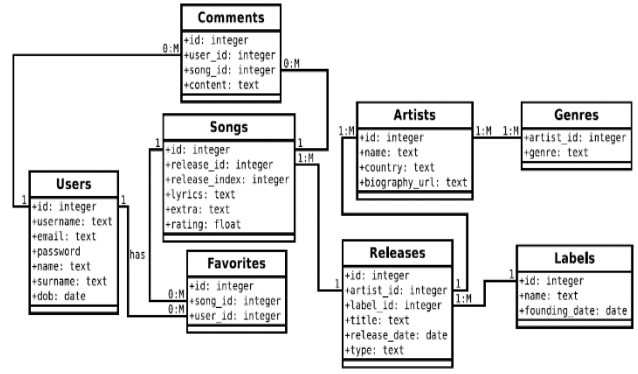
NoSQL veri tabanları göreceli olarak yeni bir gelişmedir. Fakat e-ticaret, internet arama motorları ve sosyal ağlar gibi büyük ölçekli internet uygulamaları için güvenilirliğini kanıtlamıştır. Birçok kayıt saklama teknolojisi kendilerini NoSQL ürünü olarak sınıflandırmaktadır ve her biri kendilerine özgü karakteristiklere sahiptir [18].

En bilinen lider NoSQL ürünlerinin teknik karşılaştırmaları Tablo 1'de verilmiştir [19].

Tablo 1: Lider NoSQL ürünlerinin teknik karşılaştırması
(The Technical comparison about the leader NoSQL's products)

	MongoDB	CouchDB	Riak	Redis	Voltdemort	Cassandra	HBase
Dil	C++	Erlang	Erlang	C++	Java	Java	Java
Lisans	AGPL	Apache	Apache	BSD	Apache	Apache	Apache
Model	Document	Document	Key/Value	Key/Value	Key/Value	Wide Column	Wide Column
Protokol	BSON	HTTP/REST	HTTP/REST or TCP/Protocol	bufs TCP		TCP/Thrift	HTTP/REST or
Depolama	Memory mapped b-trees	COW-BTree	Pluggable: InnoDB, LevelDB, Bitcask	In memory, snapshot to disk	Pluggable: BSV, MySQL, in-	Memtable/SSTable	HDFS

	Esinlenen						
Arama	Evet	Hayır	Evet	Hayır	Hayır	Evet	Evet
Sadeleştirme	Evet	Hayır	Evet	Hayır	Hayır	Evet	Evet
		Dynamo	Dynamo		Dynamo	BigTable, Dynamo	BigTable



Şekil 6.1 MySQL veri tabanı şeması
(Relational database schema)

6. VERİTABANI MİMARİLERİNİN PERFORMANS KARŞILAŞTIRMASI (PERFORMANCE COMPARISON OF DATABASE ARCHITECTURE)

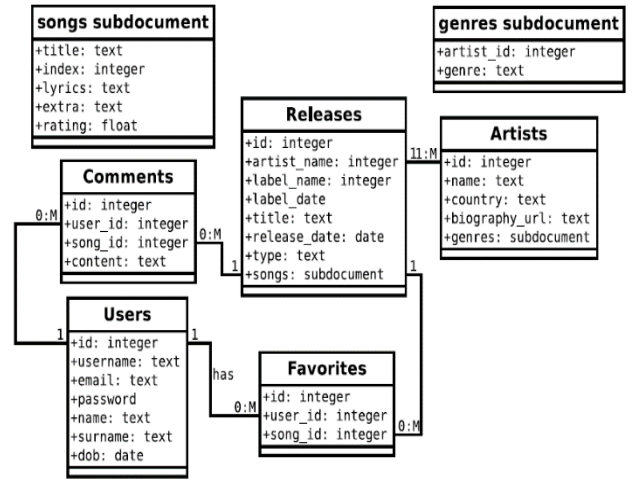
Veri tabanı mimarilerinde oldukça bol çeşit ve bir o kadar da seçenek vardır. Bu çalışmada ilişkisel veri tabanı sistemlerinden biri olan MySQL ve ilişkisel olmayan (NoSQL) veri tabanı olarak ilişkisel veri tabanı sistemlerine alternatif bir çözüm olarak ortaya çıkan, yatay olarak ölçeklendirilen bir veri depolama sistemi olan MongoDB veri tabanı sistemi kullanılmıştır.

Yapılan çalışmada; MySQL ve MongoDB veri tabanı sistemlerinin performans ve yatay ölçeklenebilirlik incelemesi için aşağıdaki işlemlerin uygulanması ve sonuçlarının ortaya çıkarılması hedeflenmiştir.

Bunlar;

- Veri tabanı sunucu sistemleri özellikleri belirlenmesi,
- Veri tabanı şemaları oluşturulması,
- Sorguların belirlenmesi,
- Veri tabanı ayarlarının yapılması,
- Ölçümler ve ölçüm metrikleri bilgileri,
- Performans analizi ve sonuçlarıdır.

Veri Tabanı Şeması: Projede iki adet veri tabanı şeması tasarlanmıştır. Biri MySQL (şekil 6.1), diğeri ise MongoDB (şekil 6.2) veri tabanıdır. Şemalar, kendi zevk ve tercihleri doğrultusunda diğer kullanıcılara şarkılar önermek için tasarlanmış farklı algoritmalar kullanan bir müzik uygulaması etrafında modellenmiştir. Tablolar arasında herhangi bir veri tekrarını ortadan kaldırmak için normalizasyon değerlendirmesi sağlanmıştır.



Şekil 6.2 MongoDB şeması
(MongoDB schema)

Veri Tabanı Sorguları: Bu çalışmada üç farklı veri tabanı sorgusu kullanılmıştır. Birinci sorgu için sadece "SELECT" deyimi içeren basit bir sorgu hazırlanmıştır. İkinci sorgu için daha karmaşık "INNER JOIN" deyimi içeren bir sorgu hazırlanmıştır. Üçüncü sorgu için ise "SELECT" ile birlikte iç içe "JOIN", "INNER JOIN" ve "WHERE" deyimi içeren detaylı karmaşık bir sorgu hazırlanmıştır.

Sorgu 1: Basit

```
SELECT * FROM Users WHERE username = 'username'
```

Sorgu 2: Karmaşık

```
SELECT 'Favourites. song_id' AS fSID, 'Favourites. user_id' AS fUID
FROM Favourites AS b INNER JOIN Favourites AS a
ON b. user_id = a. user_id
WHERE a. song_id = 123456 AND a. user_id != 987654
```

Sorgu 3: Detaylı ve karmaşık


```

SELECT ' Songs. release_id ' AS sld , ' Releases. id ' AS rld
FROM Songs INNER JOIN Releases
ON Songs. release_id = Releases. id
WHERE artist_id IN
SELECT ' Genres. artist_id ' AS gAID
FROM Genres AS c
INNER JOIN Artists AS d
ON c.artist_id = d.id WHERE d.name = ' artist_name '

```

Ölçümler: Projede ölçümler için öncelikle zaman kavramı ön planda tutulması hedeflenmiştir. Zaman ölçümleri için üç yöntem ile hareket edilmiştir.

Birinci yöntem; Clock() fonksiyonu kullanımı ile belirli bir süre CPU üzerinde harcanan zaman sonuçlarının elde edilmesini sağlamaktır.

İkinci yöntem; milisaniye hassasiyetiyle zamanlamaları sağlayan Gettimeofday() fonksiyonu kullanılarak sonuçların elde edilmesini sağlamaktır.

Üçüncü yöntem; Slow Query Log (Yavaş sorgu kaydı) olarak adlandırılmaktadır. Her veri tabanı zamanı ölçmek için kendi yöntemini sunmaktadır. Bir veri tabanı için önceden belirlenmiş uzun süren sorguları kaydedebilir ve mikro saniye doğruluğu için yapılandırılabilir [16]. Ölçüm Metrikleri: Veri tabanlarının performansını ölçmek için ortak bir metrik gereklidir. Bir uygulama için en önemli faktör, bir görevi tamamlamak için gereken süre ve veri tabanının bir işlemi tamamlaması durumu için gerekli zamandır. Bu kavramlar iyi anlaşılmalı ve birbirinden ayrı tutulmalıdır. Aşağıdaki formül sorguları hesaplamak için kullanılmaktadır.

$$\text{Saniyedeki Sorgu} = \frac{\text{Toplam Sorgu Sayısı} \times \text{Toplam İş Parçacığı Sayısı}}{\text{Ortalama Sorgu Süresi}}$$

Her iş parçacığının saniye saniye sorgu başına nasıl tepki verdiğini ölçmek için aşağıdaki formül kullanılır.

$$\text{İş Parçacığı başına saniyedeki sorgular} = \frac{\text{Toplam Sorgu Sayısı}}{\text{Ortalama Sorgu Süresi}}$$

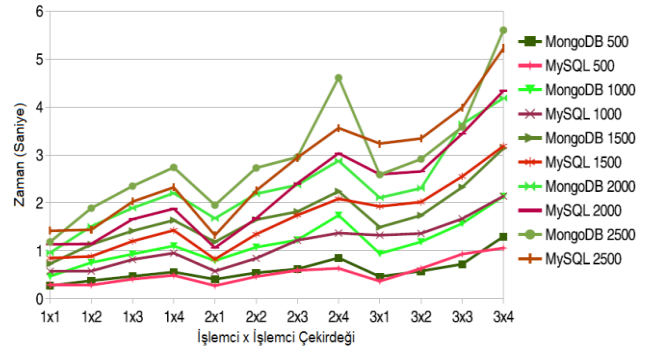
Analiz ve sonuçlar: Burada öncelikle veri tabanlarının farklı sorgu türlerine göre nasıl yanıt verdiği hem okuma hem yazma ile analiz edilen sorguların toplam sayısı ve sonuçları şekillerle gösterilmiştir. Son olarak veri tabanı boyutunun performansa etkisi konusunda inceleme yapılmıştır.

Yapılan çalışmada daha önce açıklanan koşullar kapsamında veri tabanlarının detaylı olarak karşılaştırılabilmesi için çok çeşitli durumlar yaratılmak istenmiştir. Ölçüm için kullanılan yapılandırmalar 1'den

3'e kadar işlemci sayısı ve 1'den 4'e kadar işlemci çekirdek sayısı olarak değişmektedir. Ölçümlerde yapılan sorgu sayısı 500 ile 2500 arasındadır. Her bir ölçüm beş adet test ile bitirilmiştir. Her test sonucunda sorgulardan her birini gerçekleştirmek için alınan ortalama süreler hesaplanarak raporlanmıştır.

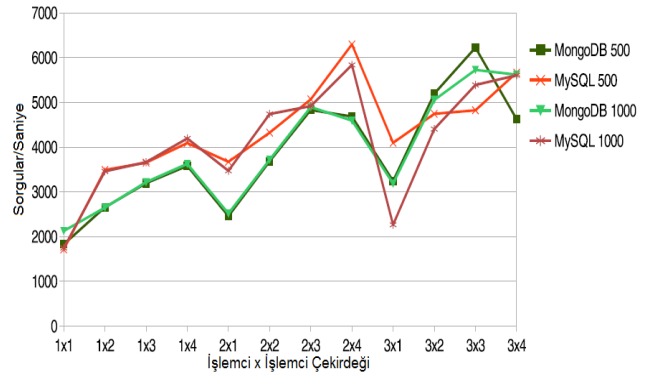
MySQL ve MongoDB veri tabanları sistemlerinin her ikisine eşdeğer miktarda sorgular yapılmaktadır. Bu alandaki yapılan sorgular veri tabanı sorguları bölümünde yer alan sorgu koşullarına göre, ölçümler ise ölçüm metriklerinde tanımlanan formüllere göre hesaplanarak yapılmaktadır.

Şekil 6.3'de MySQL ve MongoDB veri tabanlarına sorgu 1 (basit sorgu) ile karşılaştırma testi uygulanmıştır. Yapılan analizde; MongoDB, sorgu sayısı farkı arttıkça daha belirgin bir performans kötülüğü gösterdiği tespit edilmiştir. Bu karşılaştırma, işlemci çekirdeği sayılarının toplam sayısı aynı olduğu zaman, 2 ya da 1 işlemci kullanımının değişmez olduğunu açıkça ortaya koymuştur (1x2 ve 2x1). MySQL veri tabanının, özellikle 3 işlemci sayısı ile 1 işlemci çekirdeği sayısına göre incelendiğinde daha kötü performans gösterdiği görülmektedir.



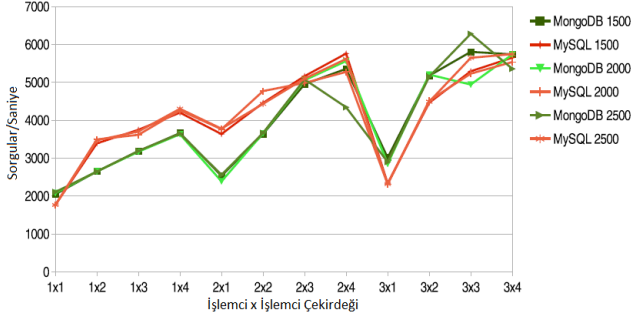
Şekil 6.3 Sorgu 1 - Analiz işlemi
(Query 1 - Analysis process)

Ayrıca, sorgular/saniye ölçüm metrik grafiği ile de şekil 6.4'de görüldüğü üzere ayrıntılı ortalama süre sonuçları elde edilmiştir.



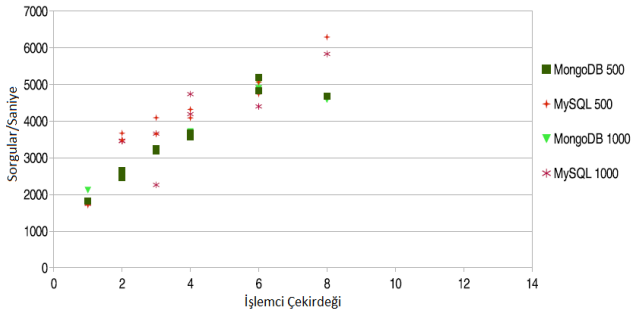
Şekil 6.4 Sorgu 1 - Sorgu/saniye analiz işlemi
(Query 1 - Queries/second analysis process)

Şekil 6.5’de MySQL veri tabanı sisteminin, sorgu sayıları arttığında MongoDB üzerinde avantaj sahibi olduğu görülmektedir. Fakat 2 işlemci ve 3 işlemci çekirdeği yapılandırmasından sonraki diğer yüksek işlemci-işlemci çekirdeği sayılarında sorgu/saniye grafiğinde keskin bir şekilde azalma görülmektedir. MongoDB bu yapılandırmalarda daha fazla avantaj göstermiştir.



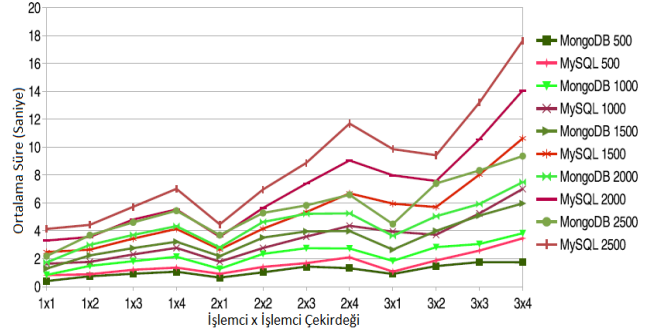
Şekil 6.5 Sorgu 1-Çok sayıda sorgu miktarı analiz işlemi
(Query 1- Many query amount analysis process)

Şekil 6.6’da işlemci çekirdeği miktarı ile saniye başına yapılan sorgu sayıları arasındaki ilişki analizi gösterilmektedir. MySQL için biraz daha iyi olan performans 4 işlemci çekirdeğine kadar hemen hemen aynıdır.



Şekil 6.6 Sorgu 1-Sorgular/Saniye ile işlemci çekirdeği miktarı için analiz işlemi
(Queries/second and amount of processor core analysis process)

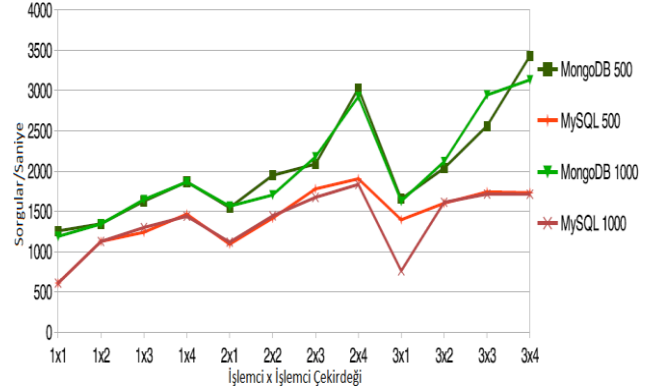
Şekil 6.7’de MySQL ve MongoDB veri tabanlarına ikinci sorgu kodu ile karşılaştırma testi uygulanmıştır. Yapılan analizde; MySQL veri tabanı sisteminin MongoDB’ye göre ortalama sorgu süreleri sonuçları, sorgu sayısı farkı arttıkça daha belirgin bir performans kötülüğü göstermiştir.



Şekil 6.7 Sorgu 2 - INNER JOIN ile karmaşık sorgu analizi işlemi

(complex query analysis process with INNER JOIN)

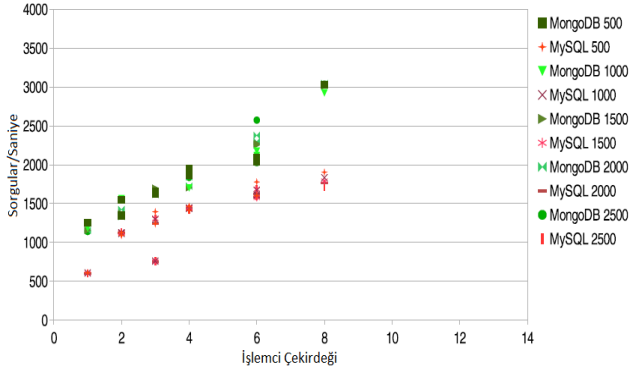
Şekil 6.8’de MySQL ve MongoDB veri tabanlarına ikinci sorgu kodu ile karşılaştırma testi uygulanmıştır. Bu test 500 ve 1000 gibi küçük veri kayıtları üzerinde yapılmıştır. Yapılan analizde; MongoDB veri tabanı sisteminin, daha az bir sürede daha çok sorgu yürütmesinin mümkün olduğu, sorgu sayısı değiştikçe performans ölçümünün daha belirgin hale gelerek sorgu/saniye başına %40 oranında daha iyi performans sergilediği gözlemlenmiştir.



Şekil 6.8 Sorgu 2- INNER JOIN ile 500 ve 1000 veri için sorgu/saniye analizi işlemi

(500 and 1000 data query / second analysis process with INNER JOIN)

İşlemci çekirdeği miktarı ile saniye başına yapılan sorgu sayıları arasındaki ilişki analizi şekil 6.9’da gösterilmektedir. İkinci sorgu kodu ile karşılaştırma testi uygulanmıştır. Bu test 500 ile 2500 veri kayıt setleri üzerinde yapılmıştır. MySQL veri tabanı sistemi, veri kayıt miktarı ve sorgu sayısı artışına göre öncelikle kademeli bir düşüş ve ardından küçük performans artışları gösterdiği tespit edilmiştir. MySQL, artan işlemci çekirdeği ve veri kayıt miktarlarında birbirine çok yakın sorgu/saniye performansı göstermiştir. MongoDB veri tabanı sisteminin, MySQL’e göre oldukça yüksek bir performans sergilediği gözlemlenmiştir. Aynı veri kayıt setlerinde MongoDB’nin MySQL üzerindeki belirgin performans farkı ve avantajı görülmektedir.

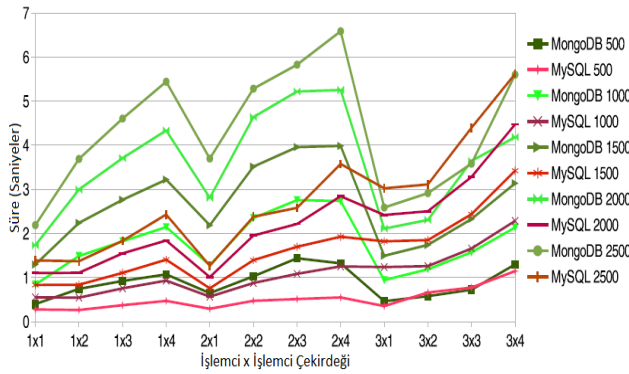


Şekil 6.9 Sorgu 2- INNER JOIN ile işlemci çekirdeği miktarı üzerinde analiz işlemi

(analysis on the amount of processor core operation with INNER JOIN)

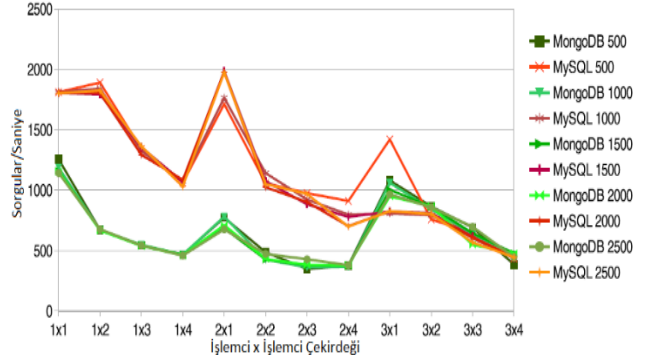
Şekil 6.10'da iç içe geçmiş "SELECT" ve "WHERE" işlemlerini içeren üçüncü sorgu neticesinde ortaya çıkan performans değerleri gösterilmektedir.

Yapılan analizlerde; MySQL veri tabanı sisteminin MongoDB 'ye göre sorgu süresi sonuçları, veri kayıt sayısı farkı arttıkça iyi bir performans göstermiştir. Fakat işlemci ve işlemci çekirdeği yapılandırılmalarının değiştirildiği durumlarda performans farklılıkları daha belirgin hale gelmiştir. Bu karşılaştırma, işlemci ve işlemci çekirdeği sayılarının 3x1, 3x2, 3x3 ve 3x4 şeklinde yapılandırıldığı anlarda iki veri tabanı birbiri üzerine hemen hemen aynı performansı gösterdiği gözlemlenmiştir.



Şekil 6.10 Sorgu 3 - Detaylı karmaşık sorgu süre analizi
(Detailed and complex query time analysis)

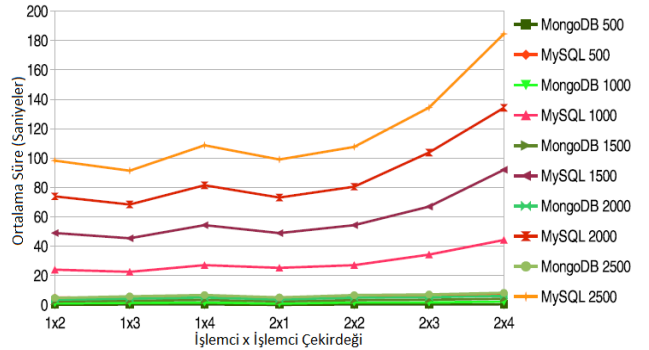
MySQL ve MongoDB veri tabanlarına üçüncü sorgu kullanılarak uygulanan karşılaştırma testi sonuçları Şekil 6.11'de daha iyi anlaşılabilir. MySQL veri tabanı sisteminin 2x4 işlemci ve işlemci çekirdeği yapılandırmasında en iyi performansı gösterdiği açık bir şekilde görülmektedir. Fakat 2x1 ve 3x1 işlemci ve işlemci çekirdeği yapılandırmalarında her iki veri tabanı için performans sorunu olduğu gözlemlenmektedir. Bu performans sorunu MySQL'de daha belirgin haldedir.



Şekil 6.11 Sorgu 3- Detaylı ve karmaşık sorgu ile Sorgular/saniye analiz işlemi

(Queries/second analysis process with detailed and complex queries)

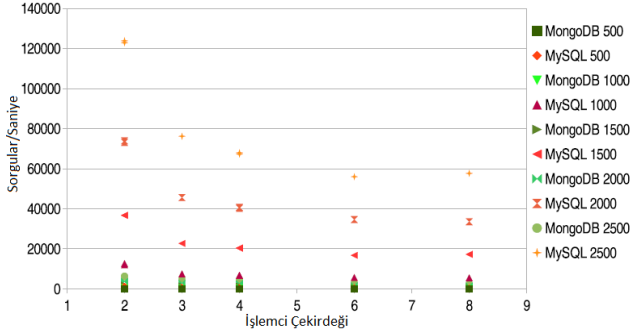
Şekil 6.12'de üçüncü sorgu ile ortalama süre ölçümleri gösterilmiştir. Yapılan analizde; MySQL veri tabanı sisteminin MongoDB'ye göre ortalama sorgu süreleri sonuçları, veri kayıt sayısı farkı arttıkça oldukça belirgin bir performans kötülüğü gösterdiği gözlemlenmiştir. Bu karşılaştırma, eşdeğer veri kayıt setlerinde yapılmasına rağmen MySQL veri tabanının sorguları işleme ve sonuçlandırma süresi ortalamasının oldukça yüksek olduğu görülmektedir. MongoDB daha belirgin ve istikrarlı bir performans göstermektedir.



Şekil 6.12 Sorgu 3 - Detaylı ve karmaşık sorgu kodu ile ortalama süre analiz işlemi

(The average time analysis process with detailed and complex query code)

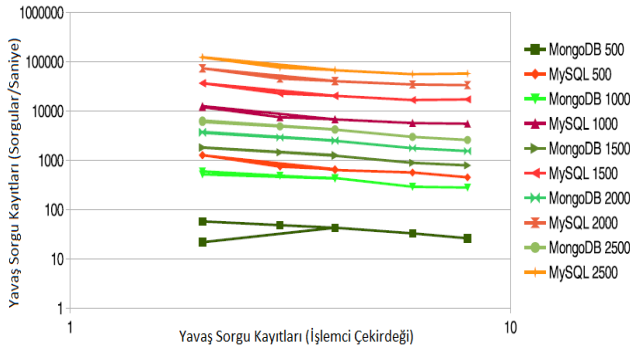
MySQL ve MongoDB veri tabanlarına üçüncü sorgu olarak tanımlanan detaylı ve karmaşık sorgu kodu içeren karşılaştırma testi analizi Şekil 6.13'de gösterilmiştir. Bu test 500 ile 2500 veri kayıt setleri üzerinde yapılmıştır. MySQL veri tabanı sistemi, iki eksen boyunca logaritma kullanılarak çizilen grafikte logaritmik bir eğilim olduğu görüntüsü sergilemektedir. MongoDB'nin ise eğilimi nerede ve nasıl gösterdiği net olarak görülmemektedir.



Şekil 6.13 Sorgu 3 - Detaylı ve karmaşık sorgu ile işlemci çekirdeği üzerinde analiz işlemi

(Process analysis of processor cores with detailed and complex queries)

Zamanlama ölçeği büyütülerek veri tabanları sistemleri arasındaki performans farkının Şekil 6.14’de daha anlaşılabilir hale geldiği görülmektedir. Ölçek büyüdükçe MySQL’in performansındaki dezavantaj açıkça görülmektedir. MongoDB tüm veri kayıt setlerinde oldukça iyi bir performans gösterdiği ortaya konulmuştur.



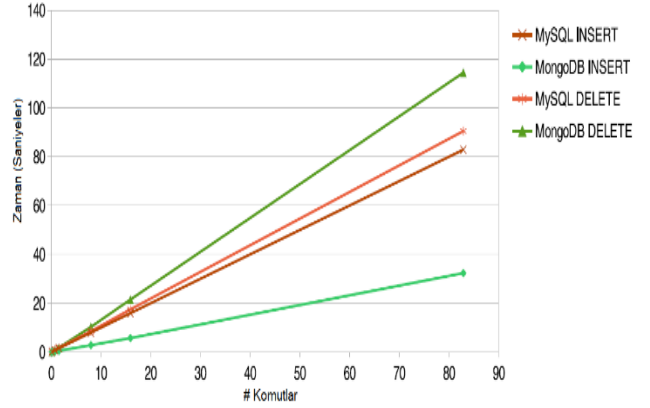
Şekil 6.14 Sorgu 3- Detaylı ve karmaşık sorgu ile ölçeklendirilmiş analiz işlemi

(Scaled analysis process detailed and complex queries)

Son olarak MySQL ve MongoDB veri tabanlarına veri ekleme “INSERT” ve silme “DELETE” işlemleri uygulanmıştır. Veri tabanı komutlarının çoğu veri okuma ve yazma sorguları olmasına rağmen silme sorguları da önemli bir faktör olarak görülmektedir. Şekil 6.15’de her iki veri tabanı sisteminin INSERT ve DELETE işlemlerine ait performans grafiği gösterilmektedir.

Yapılan analizde; her iki veri tabanının komut sayılarına göre işlem süreleri doğrusal bir eğilim göstermektedir. MongoDB’nin veri ekleme işlemi MySQL’e göre çok daha iyi bir performansa sahiptir.

Veri silme işleminde ise MongoDB’nin MySQL ile benzer bir performansa sahip olduğu fakat veri silme komut sayılarının artışı ile MySQL veri tabanı sisteminin silme işleminde iyi bir performans sergilediği gözlemlenmiştir.



Şekil 6.15 INSERT ve DELETE işlemleri

(INSERT and DELETE operations)

7. SONUÇ VE DEĞERLENDİRME (RESULT AND EVALUATION)

Klasik ilişkisel veri tabanlarını altkümü olarak gören, aynı zamanda SQL ve Daha Fazlası (Not Only SQL) olarak da adlandırılan dağıtık mimari ile oluşturulmuş veri tabanları ile ilişkisel veri tabanları karşılaştırılmış ve yönetim bilişim sistemleri açısından incelenmiştir.

Bu çalışmada, yönetim bilişim sistemleri kapsamında veri tabanlarının modellenmesi ve niteliklerinin belirlenmesi, en uygun performans ölçümleri, sürecin uygun hale getirilmesi ve en uygun veri tabanı seçiminde kullanıcılara ışık tutulması hedeflenmiştir. Son yıllardaki teknolojik ilerleme ile paralel olarak yapılan çalışmalar, veri tabanı sistemlerinde de etkisini göstererek NoSQL kavramını gün yüzüne çıkartmıştır. Hazırlanan bu makalede ilişkisel ve ilişkisel olmayan veri tabanı yönetim sistemlerinin performans karşılaştırmasının yapılması ve farklı faktörlerin her bir veri tabanını nasıl etkilediğini keşfederek hangi teknolojinin hangi durumlarda diğerinden daha uygun olduğunun araştırılması amaçlanmıştır.

Yapılan literatür taramasında en yakın olarak görülen çalışmalarda; Ahmet ALADILY tarafından 2015 yılında yazılan “En çok kullanılan NoSQL veri tabanları performans karşılaştırması” konulu tezde 4 ayrı dağıtık veri tabanı (Couchdb, Mongoddb, Cassandra ve Hbase) performansı karşılaştırıldığı görülmüştür. Yılmaz GÖKŞEN ve Hakan AŞAN tarafından hazırlanan “Veri büyüklüklerinin veri tabanı yönetim sistemlerinde meydana getirdiği değişim: NoSQL” konulu çalışmada NoSQL veri tabanları özelinde bilgiler verilerek esneklik, tutarlılık, performans, sorgulama dili ve veri bağlantıları kapsamında ilişkisel veri tabanları ile genel bir karşılaştırma yapılmıştır. Bu çalışmada ise yapılan uygulama ile bilinen tutarlılık, esneklik gibi özelliklerden bahsedilmek sureti ile performans özelinde çok farklı durumlar yaratılarak detaylı bir analiz yapılmıştır.

Artan yazılım rekabetinde kullanıcının en önemli tercih kısıtlarından biri olan çalışma hızı dikkate alınarak, oldukça yaygın kullanım alanına sahip veri tabanı yönetim sistemlerinden MongoDB ve MySQL'in mümkün olduğunca eşit koşullarda işlem süreleri hesaplanarak performansları karşılaştırılmıştır. Farklı sorgu tipleri çalıştırılan testlerde, detaylı ve karmaşık yapılandırmalar ile veri tabanları analiz edilmiştir.

Yapılan analizlerde NoSQL ağırlıklı bir veri tabanının büyük miktarda veri çiftleri içerebildiği, veri çoğaltmada da basit şeması nedeniyle MongoDB kullanılarak daha hızlı daha karmaşık sorgu tiplerinin çalıştırılabildiği izlenmiştir. Her iki veri tabanı sisteminde farklı yapılandırma durumlarında ikinci sorgu tipi ile yapılan performans testlerinde, MongoDB veri tabanı sistemi MySQL'e göre en iyi performansı göstermiştir. Sonraki karşılaştırma testlerinde detaylı ve karmaşık sorgular ele alınarak MongoDB, alt belge koleksiyonu kullanımı nedeniyle MySQL üzerinde çok büyük bir avantaja sahip olduğunu göstermiştir. Bu avantaj, veri tekrarı yaşanabilme durumu pahasına dikkate değer bir şekilde görülmüştür. Bu tür sorgularda büyük veri tabanı boyutundan kaynaklanan depolama ve bellek miktarı maliyetini hesaplarken alternatif olarak NoSQL veri tabanlarını dikkate almak çok önemlidir.

Yapılan son performans testleri ise yazma ve silme işlemleridir. Basit arama sorgularında mantıksal olarak veri silme işlemi dikkate alınarak yapılan karşılaştırmalar sonucunda MySQL veri tabanı sistemi iyi bir performans göstermiştir. Öncelikle silinecek verinin bulunması gerektiğinden silme işlemine direk olarak bağlantılanır. Her iki veri tabanı bu karşılaştırma sonucunda doğrusal bir eğilim gösterirken MongoDB eklemeler sırasında MySQL'e göre oldukça belirgin ve çok daha iyi bir performans göstermiştir.

Yapılan testlerin bir diğer önemli yönü, işlemci ve işlemci çekirdeklerinin farklı şekillerde yapılandırılarak kullanılmasıdır. Bu çalışmada, veri tabanlarının 1, 2 ve 3 işlemci ile birden fazla işlemci çekirdeği üzerinde nasıl performans göstereceğinin test edilmesi ve karşılaştırılması için çoklu sorgulama işlemleri kullanılmıştır. Bu test detaylı ve karmaşık sorgular ile birlikte yapılmıştır. Burada sorgu karmaşıklığına bağlı olarak veri tabanları farklı davranmaktadır. Bu farklılık ise daha fazla sorgu sayısı ile sorgular/saniye analiz grafiklerinde görülmüştür.

Çok çeşitli durumlar yaratılarak elde edilen sonuçlar ile işletmelere hangi durumda nasıl bir veri tabanı yönetim sistemi kullanmaları konusunda fikir verilmiştir.

Sonuç olarak, farklı kriterler ile bu veri tabanlarını incelediğimizde iki veri tabanının da avantaj ve

dezavantajları olduğu görülmüştür. İlişkisel veri tabanı yönetim sistemlerinin kullanıldığı uygulamaların ilişkisel olmayan (NoSQL) sistemlere taşınmasının ilk etapta zor olması, veri kaybının söz konusu olabilmesi ve NoSQL veri tabanı sistemlerinin veri güvenliği alanında ilişkisel veri tabanı yönetim sistemleri kadar mesafe kat etmemiş olması gibi dezavantajları olsa dahi hız, geliştirme zamanı ve ölçeklenebilirlik gibi özellikleri ile ilişkisel olmayan (NoSQL) veri tabanlarının kullanılması performans açısından daha etkin sonuçlar almamızı sağlayacaktır.

KAYNAKLAR (REFERENCES)

- [1] A. Emhan, "Karar Verme Süreci ve Bu Süreçte Bilişim Sistemlerinin Kullanılması", Elektronik Sosyal Bilimler Dergisi, 212-224, 2007.
- [2] Y. Gökşen, "Veri Büyüklüklerinin Veri tabanı Yönetim Sistemlerinde Meydana Getirdiği Değişim: NoSQL", Bilişim Teknolojileri Dergisi, 8-3, 2015
- [3] D. Karahoca, "Yönetim Bilişim Sistemleri", Beta Basım Yayım Dağıtım A.Ş., İstanbul, 1998.
- [4] İnternet: E. Baycan, "Veri Tabanı Yönetim Sistemleri", <http://www.gencklavyeler.com/forum/attachment.php?aid=1135>.
- [5] İnternet: "Veri Tabanı Modeli", http://en.wikipedia.org/wiki/Database_model, 6.02.2017.
- [6] E. Önder, "Yönetim Bilişim Sistemleri Kapsamında Web Tabanlı İlişkisel Veri tabanı Yönetim Sistemleri ve Bir Uygulama", Yüksek Lisans Tezi, İstanbul Üniversitesi, Sosyal Bilimler Enstitüsü, İşletme Anabilim Dalı, Sayısal Yöntemler Bilim Dalı, 2005.
- [7] Ö. Özasan, "Web Tabanlı Jeodezik Amaçlı Mekânsal Veri Tabanı tasarımı ve Uygulamaları" Doktora Tezi, İstanbul Teknik Üniversitesi, Fen Bilimleri Enstitüsü, Jeodezi ve Fotogrametri Mühendisliği, Geomatik Mühendisliği, 2011.
- [8] E.F. Codd, "A Relational Model of Data for Large Shared Data Banks", 1970.
- [9] D. Gündüz, "Veri tabanlarına Giriş Seminer Notları", IV. Akademik Bilişim Konferansı, Konya, 2002.
- [10] Ö. Özasan, "Jeodezik Veri Tabanı Tasarımı ve WEB Tabanlı Yönetimi" Yüksek Lisans Tezi, İstanbul Teknik Üniversitesi, Fen Bilimleri Enstitüsü, 2003.
- [11] R. Stephens, R. Plew, "Veri Tabanları", Çeviren: Nalan Güven Küçükler, Alfa Basım Yayım Dağıtım Ltd. Şti, İstanbul, 2003.
- [12] F. Sürmeli, M. Erdoğan, N. Erdoğan, K. Banar, E. Kaya, A. Sevim, "Muhasebe Bilgi Sistemi", Anadolu Üniversitesi, Eskişehir, 2006.
- [13] İnternet: "Acid" <http://en.wikipedia.org/wiki/Acid>; 08.08.2014.
- [14] İnternet: S. Davaz, "NoSQL Nedir Avantajları ve Dezavantajları Hakkında Bilgi" <https://blog.kodcu.com/2014/03/nosql-nedir-avantajlari-ve-dezavantajlari-hakkinda-bilgi>, 28.03.2014.
- [15] G. Vaish, "Getting Started With NoSQL", Packt Publishing, United Kingdom, 2013.
- [16] M. Otey, "NoSQL? No Way!", SQL Server Magazine, pp:5, 2010.
- [17] İnternet: "Oracle NoSQL Database", <http://www.oracle.com/technetwork/database/nosqldb/learnmore/nosql-database-498041.pdf>, 11. 2011.
- [18] N. Rozanski, E. Woods, "Software Systems Architecture: Working with Stakeholders Using Viewpoints and Perspectives", Pearson Education Inc., USA, 2012.
- [19] G. Burd, "NoSQL", Sysadmin Journal, Vol:36 No. 5 ss:5-12 ABD, 2011.

Attribute Reduction in Stochastic Information Systems Based on α -Dominance

Emel KIZILKAYA AYDOĞAN, Mihrimah ÖZMEN

Department of industrial engineering, Erciyes University, TURKEY

emelkizilkaya@gmail.com, mihrimah@erciyes.edu.tr

(Geliş/Received:03.11.2016; Kabul/Accepted:22.01.2017)

DOI: 10.17671/gazibtd.309305

Abstract— Rough set has been commonly taken part in literature to examine inadequate and incomplete information systems. The efficiency of rough set with stochastic data observed for developing convenience and scalability. In this study, we use a ranking approach for attribute reduction in stochastic information systems and generalized this via presenting a dominance relation. We obtained the rough set approach of attribute reduction in stochastic information systems by establishing the dominance degrees. Furthermore, attribute reduction methods are studied by considering discernibility matrix and this approach is applied to explanatory examples to demonstrate its validity. Also this research proposes many research fields and new application areas show a tendency to concerning rough set approach to stochastic information systems.

Keywords- *Rough Sets, Stochastic Information Systems, Attribute Reduction, Dominance Relation*

1. INTRODUCTION

As a soft computing technique, rough set theory is proposed by Pawlak and it is an important theory to deal with insufficient information [1,2]. The theory is an expanded of the classical set theory for modelling uncertainty or imprecision information. It has conceived as a powerful mathematical tool for knowledge discovery. Particularly it can be used in evaluation the significance of attributes and derive decision rules and also to describe the dependencies among attributes [2,3]. It is useful in many fields such as machine learning, data mining and pattern recognition [4-8]. The information systems, the basis of rough set based data analysis; contain data about objects of interest, characterized by a finite set of attributes [9,10].

The classical rough set theory does not examine the problem of ordering objects. However, note that indiscernibility-based rough set model is powerless in many real situations that are the ordering of objects attributes with preference-ordered domains. To solve this problem, Greco, Matarazzo, and Slowinski have given to the literature, the dominance-based rough set approach (DRSA) which is an extension of rough set theory [11-14]. DRSA considers the ordering characteristics of attributes. Many researches have been studied about DRSA, lately [15-17].

Stochastic data is to define objects with uncertain judgements. It is also influenced by varied kinds of errors such as measurement, computation errors etc. Also stochastic information systems ensures us a powerful tool for data analyses and they are found almost everywhere in real world applications. Researchers from all over the world interested in stochastic data from among so many complex data. So the stochastic extension of DRSA was generalized as the notion

of lower approximations to the stochastic case by estimating the class intervals for each object [18].

Nowadays, researchers are proposed approaches using stochastic dominance (SD) rules [19-28]. Some of these methods are to rank alternatives by using the determined SD relations based on rough set theory [26-27]. Moreover, the approach depends on stochastic dominance degree (SDD) has been also proposed [28].

The aim of this paper is to future researches rough set in stochastic information systems based on dominance relation. Nevertheless, to solve applicable problem, just only one knowledge granulation may not be fine enough. For this reason the issues of multi granulation have been absolutely discussed in many rough set studies [29-34]. From this aspect, we will propose a new dominance approach into stochastic information systems. This mechanism is to apply a parameter α , which checks the size of dominance-based information granulation.

The remaining of this study is consisted of as follows. Dominance relation in stochastic information systems are shortly reminded in Section 2. In Section 3, attribute reduction in stochastic information systems based on discernibility matrix is introduced and some important properties are discussed. In Section4, illustrative examples for the two more cases are investigated to show the possible implementation of the proposed approach. And lastly, Section 5 summarizes and emphasizes the main properties of this study.

2. α -DOMINANCE RELATION IN STOCHASTIC INFORMATION SYSTEMS

In this section we present a dominance relation to a stochastic information system by using SDD. Stochastic dominance is a comparing technique for uncertain alternatives in decision making [36]. And also SDD was presented as an approach to solve the stochastic multiple criteria decision making problem, where results of alternatives according to criteria are demonstrated by random variables with probability distributions by Liu et. al [35].

In decision making, two kinds of problems are examined: to obtain ranking based on information aggregation, and to obtain dominance rules with relations. In this study, we only work on to rank all objects with the dominance relation in a stochastic information system. In their study, Zhang and Qiu describe a dominance degree notion to rank all objects in classical ordered information systems [37]. Hereinafter, we define a dominance degree between two objects and a whole dominance degree of an object.

A stochastic information system (SIS) is a quadruple $S = (U, AT, V, D)$ where U is a finite non-empty set of objects and AT is a finite non-empty set of attributes, $V = \cup_{a \in AT} V_a$ and V_a is a domain of attribute a , $D: f_1(x) \times f_2(x) \times AT \rightarrow V$ is a total function such that $f_1(x)$ and $f_2(x)$ are probability distributions of two objects x_1 and x_2 respectively, $D(f_1(x), f_2(x), a) \in V_a$ for every $a \in AT, x_1$ and $x_2 \in U$ called an information function where V_a is a set of stochastic data.

In comparing two probability distributions, there are two kinds of random variables, i.e. continuous (probability density functions) and discrete (probability mass functions) random variables. So, two probability distributions are compared with three possible states. i.e., (1) two continuous probability distributions, (2) two discrete probability distributions and (3) a continuous probability distribution and a discrete probability distribution [35]. Dominance degree between two objects with respect to the dominance relation is defined as:

Definition 1 (Stochastic Dominance Relation Between Two Continuous Probability Distribution)

Let X_1 and X_2 be two independent continuous random variables with probability distributions $f_1(x)$ and $f_2(x)$, respectively, where $\int_{-\infty}^{+\infty} f_1(x)dx = 1$ and $\int_{-\infty}^{+\infty} f_2(x)dx = 1$. Then the dominance degree of $f_1(x)$ over $f_2(x)$ (noted as $D_{f_1 > f_2}$) is given by

$$D_{f_1 > f_2} = \int_{-\infty}^{+\infty} \int_{-\infty}^{x_1} f_1(x_1) f_2(x_2) dx_2 dx_1, \tag{1}$$

and accordingly, the dominance degree of $f_2(x)$ over $f_1(x)$ (noted as $D_{f_2 > f_1}$) is given by

where $D_{f_1 > f_2} + D_{f_2 > f_1} = 1$,

$$D_{f_2 > f_1} = \int_{-\infty}^{+\infty} \int_{x_1}^{+\infty} f_1(x_1) f_2(x_2) dx_2 dx_1. \tag{2}$$

Definition 2 (Stochastic Dominance Relation Between Two Discrete Probability Distribution)

Let Y_1 and Y_2 be two independent discrete random variables with probability distributions $g_1(y)$ and $g_2(y)$, respectively, where $\sum_{y=-\infty}^{+\infty} g_1(y) = 1$ and $\sum_{y=-\infty}^{+\infty} g_2(y) = 1$. Then the dominance degree of $g_1(y)$ over $g_2(y)$ (noted as $D_{g_1 > g_2}$) is given by

$$D_{g_1 > g_2} = \sum_{y_1=-\infty}^{+\infty} \sum_{y_2=-\infty}^{y_1} g_1(y_1) g_2(y_2) - 0.5 \sum_{y_1=-\infty}^{+\infty} g_1(y_1) g_2(y_1), \tag{3}$$

and accordingly, the dominance degree of $g_2(y)$ over $g_1(y)$ (noted as $D_{g_2 > g_1}$) is given by

where $D_{g_1 > g_2} + D_{g_2 > g_1} = 1$,

$$D_{g_2 > g_1} = \sum_{y_1=-\infty}^{+\infty} \sum_{y_2=y_1}^{+\infty} g_1(y_1) g_2(y_2) - 0.5 \sum_{y_1=-\infty}^{+\infty} g_1(y_1) g_2(y_1). \tag{4}$$

Definition 3 (Stochastic Dominance Relation Between Continuous Probability Distribution and Discrete Probability Distribution)

Let X be an independent continuous random variable with probability distribution $f(x)$, and Y be an independent discrete random variable with probability distribution $g(y)$, where $\int_{-\infty}^{+\infty} f(x)dx = 1$ and $\sum_{y=-\infty}^{+\infty} g(y) = 1$. Then the dominance degree of $f(x)$ over $g(y)$ (noted as $D_{f > g}$) is given by

$$D_{f > g} = \sum_{y=-\infty}^{+\infty} \left[g(y) \int_y^{+\infty} f(x) dx \right], \tag{5}$$

and accordingly, the dominance degree of $g(y)$ over $f(x)$ (noted as $D_{g > f}$) is given by

where $D_{f > g} + D_{g > f} = 1$,

$$D_{g > f} = \sum_{y=-\infty}^{+\infty} \left[g(y) \int_{-\infty}^y f(x) dx \right]. \tag{6}$$

Definition 4 (Stochastic Dominance Relation Between Two Normal Probability Distributions)

The normal probability distribution is used extensively in that it can well model the additive effect of many independent factors [38-42]. In the following, a case that continuous probability distributions are normal ones was analyzed [35].

Let Z_1 and Z_2 be two independent normal random variables with probability distributions $h_1(z) = \frac{1}{\sigma_1 \sqrt{2\pi}} e^{-(z-\mu_1)^2 / 2\sigma_1^2}$ and $h_2(z) =$

$\frac{1}{\sigma_2\sqrt{2\pi}} e^{-(z-\mu_2)^2/2\sigma_2^2}$ respectively, where μ_1 and μ_2 are means, and σ_1 and σ_2 are standard deviations. Then the dominance degree of $h_1(z)$ over $h_2(z)$ (noted as $D_{h_1>h_2}$) is given by

$$D_{h_1>h_2} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\alpha} e^{-t^2/2} dt, \tag{7}$$

where $\alpha = \frac{\mu_1 - \mu_2}{\sqrt{\sigma_1^2 + \sigma_2^2}}$

Remark 1. Consider arbitrary two objects A_1 and A_2 in decision analysis. Let $f_1(x)$ and $f_2(x)$ be probability distributions on consequences of A_1 and A_2 , respectively. $D_{f_1>f_2}$ can be also regarded as the dominance degree of A_1 over A_2 (noted as $D_{A_1>A_2}$), i.e., $D_{f_1>f_2} \Leftrightarrow D_{A_1>A_2}$. The greater $D_{f_1>f_2}$ is, the greater the dominance degree of object A_1 over A_2 will be.

By Equations (1), (2), (3), (4), (5), (6) and (7), the dominance degree matrix D_j of object pairwise comparisons with respect to criterion a_j can be constructed, i.e.,

$$D_j = [D_{ikj}]_{m \times m} = \begin{bmatrix} D_{11j} & D_{12j} & \dots & D_{1mj} \\ D_{21j} & D_{22j} & \dots & D_{2mj} \\ \vdots & \vdots & \ddots & \vdots \\ D_{m1j} & D_{m2j} & \dots & D_{mmj} \end{bmatrix},$$

$j = 1, 2, \dots, n,$

where D_{ikj} denotes the dominance degree of x_i over x_k with respect to attribute a_j , and $D_{ikj} + D_{kij} = 1$.

First, overall dominance degree matrix D of object pairwise comparisons is constructed, i.e.,

$$D = [D_{ik}]_{m \times m} = \begin{bmatrix} D_{11} & D_{12} & \dots & D_{1m} \\ D_{21} & D_{22} & \dots & D_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ D_{m1} & D_{m2} & \dots & D_{mm} \end{bmatrix},$$

where D_{ik} denotes the overall dominance degree of x_i over x_k and is calculated by

$$D_{ik} = \frac{\sum_{j=1}^n D_{ikj}}{|AT|}, \quad i, k = 1, 2, \dots, m. \tag{8}$$

From Eq. (8), it can be seen that $D_{ik} + D_{ki} = 1$ and $0 \leq D_{ik} \leq 1$.

Let $(x_i, x_k) \in U \times U$, thus a dominance relation matrix with respect to AT can be constructed with the dominance relation. Also the whole dominance degree of each object can be obtained from this matrix with this formula

$$D(i) = \frac{1}{|U|-1} \sum_{i \neq k} D_{ik}, \quad x_i, x_k \in U. \tag{9}$$

According to these two concepts, a ranking approach for all objects can also be generated. Through the number of D_i , whole dominance degree of each object on the universe, all objects can be rank. Therefore larger value indicates that an object better. This concept can be understood with the next example.

In the following, we will give an example for the comparing of two discrete probability distributions.

Example 1.1

This example is adapted from Zaras and Martel (1994) and Liu et al. (2011) [24,35]. A stochastic information system is given in Table 1, where $U = \{x_1, x_2, \dots, x_{10}\}$, $AT = \{a_1, a_2, a_3, a_4\}$ and there are 7 decision makers.

Firstly the types of stochastic dominance relations are defined for each pair of objects according to each attribute. The larger values of attributes are preferable one than the smaller ones.

The evaluations of objects with respect to attributes are defined in the form of probability distributions (Table 1). For example, three experts in the seven give their considerations on project x_j with respect to criterion a_1 using score 2, then the ‘probability’ that the consideration on project x_j is score 2 is regarded as 3/7 (see Table 1). Firstly in Table 1, the probability distribution of random variable on each object according to each attribute is obtained. Then, the dominance degree of one object over another according to each attribute can be determined using Equation (3) and (4). So, four dominance degree matrices D_1, D_2, D_3 and D_4 are respectively established, i.e.,

Table 1. The evaluations provided by experts

Attribute	Values	Objects									
		x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}
a_1	1	0	0	0	1/7	0	1/7	1/7	1/7	0	0
	2	3/7	1/7	0	0	0	0	0	2/7	0	1/7
	3	1/7	0	0	0	1/7	0	0	2/7	0	2/7
	4	0	2/7	0	0	0	0	0	1/7	0	2/7
	5	2/7	1/7	3/7	1/7	0	0	3/7	1/7	2/7	1/7
	6	0	2/7	1/7	0	2/7	0	1/7	0	1/7	0
	7	1/7	0	1/7	0	2/7	1/7	0	0	3/7	1/7
	8	0	1/7	2/7	1/7	0	3/7	1/7	0	1/7	0
	9	0	0	0	4/7	2/7	0	0	0	0	0
	10	0	0	0	0	0	2/7	1/7	0	0	0
a_2	1	0	1/7	1/7	0	0	0	1/7	3/7	0	0
	2	2/7	0	0	0	0	0	3/7	3/7	0	1/7
	3	1/7	0	0	1/7	0	4/7	1/7	0	1/7	0
	4	0	0	0	1/7	0	0	0	1/7	1/7	0
	5	2/7	0	0	0	1/7	0	1/7	0	0	0
	6	0	1/7	1/7	1/7	2/7	0	1/7	0	1/7	0
	7	0	1/7	0	0	1/7	1/7	0	0	4/7	2/7
	8	1/7	1/7	2/7	3/7	2/7	2/7	0	0	0	3/7
	9	1/7	3/7	1/7	1/7	1/7	0	0	0	0	0
	10	0	0	2/7	0	0	0	0	0	0	1/7
a_3	1	0	0	1/7	0	1/7	0	0	2/7	0	1/7
	2	0	0	0	0	0	0	3/7	1/7	0	2/7
	3	1/7	0	0	1/7	0	0	1/7	4/7	1/7	0
	4	3/7	0	0	0	0	1/7	1/7	0	2/7	0
	5	0	1/7	0	0	0	1/7	2/7	0	2/7	0
	6	1/7	0	0	0	0	0	0	0	0	2/7
	7	0	1/7	0	1/7	0	0	0	0	2/7	2/7
	8	1/7	2/7	0	2/7	3/7	2/7	0	0	0	0
	9	1/7	3/7	2/7	1/7	1/7	1/7	0	0	0	0

Take for instance x_2 , under the given dominance degree $\alpha = 0.6$, objects x_1, x_2, x_7, x_8, x_9 and x_{10} are regarded as dominated by x_2 with respect to set of attributes AT .

3. ATTRIBUTE REDUCTION IN STOCHASTIC INFORMATION SYSTEMS BASED ON DISCERNIBILITY MATRIX

Searching significant subsets of attributes which ensure the identical or more meaning for the whole set of attributes, is one of the main research subject. These subsets are named reducts. Nowadays, many attribute reduction approaches have been studied in literature for different dominance relations. For instance, Shao and Zhang proposed an expansion of reduct dominance relation for incomplete information systems [45]. Yang et al. proposed a similarity dominance relation to form all reducts [46].

In this this part, a framework of attribute reduction with respect to α -dominance relation in stochastic information systems is established and an example is employed to demonstrate its effectiveness.

Definition 6. Let $S = (U, AT, V, D)$ be a stochastic information system, $\forall A \subseteq AT$ and $\alpha \in [0,1]$,

1. if $DOM_A^\alpha = DOM_{AT}^\alpha$, then A is referred to as a α -dominance consistent set in S ;
2. if A is a α -dominance consistent set in S and $A - \{a\}$ is not a α -dominance consistent set for each $a \in A$, i.e.,

$DOM_{A-\{a\}}^\alpha \neq DOM_{AT}^\alpha$, then A is referred to as a α -dominance reduct in S ;

3. the set of all α -dominance reducts in S is denoted by $Red^\alpha(S)$;

4. the intersection of all α -dominance reducts in S is referred to as the α -dominance core.

A α -dominance consistent set in S is a subset of full attributes which preserves α -dominance relation DOM_{AT}^α , while α -dominance reduct in S is a *minimal* α -dominance consistent set which preserves α -dominance relation DOM_{AT}^α .

In the following, the discernibility matrix based approach to compute all α -dominance reducts in S is presented.

$\forall x_i, x_k \in U$, let us denote

$$D_{AT}^\alpha(x_i, x_k) = \begin{cases} \{a \in AT: (x_i, x_k) \notin DOM_a^\alpha\} & (x_i, x_k) \notin DOM_{AT}^\alpha \\ \emptyset & \text{other wise} \end{cases}$$

$D_{AT}^\alpha(x_i, x_k)$ is showed to as the α -discernibility set for objects pair (x_i, x_k) for the given dominance degree α . Describe, furthermore, the α -dominance matrix such that

$$M^\alpha = \{D_{AT}^\alpha(x_i, x_k): \forall x_i, x_k \in U\}, \tag{14}$$

and let

$$M_0^\alpha = \{D_{AT}^\alpha(x_i, x_k): \forall D_{AT}^\alpha(x_i, x_k) \in M^\alpha, D_{AT}^\alpha(x_i, x_k) \neq \emptyset\}, \tag{15}$$

then to obtain a α -dominance consistent set according to α -dominance matrix, the following theorem can be used

Theorem 1. Let S be a stochastic information system, $\forall A \subseteq AT$, we have

$$DOM_A^\alpha = DOM_{AT}^\alpha \Leftrightarrow A \cap D_{AT}^\alpha(x_i, x_k) \neq \emptyset, \forall D_{AT}^\alpha(x_i, x_k) \in M_0^\alpha.$$

Theorem 1 ensures that adequate and essential conditions to evaluate whether a subset of attributes is a α -dominance consistent set in S .

Definition 7. Let S be a stochastic information system, let us define

$$\Delta = \bigwedge_{D_{AT}^\alpha(x_i, x_k) \in M_0^\alpha} \bigvee D_{AT}^\alpha(x_i, x_k), \tag{16}$$

Δ is referred to as the α -dominance discernibility function in S .

Theorem 2. Let S be a stochastic information system, $\forall A \subseteq AT$, A is a α -dominance reduct in S if and only if $\forall A$ is a prime implicant of Δ .

According to those described above, we can see that it ensures a practical framework for attribute reduction, which can be expressed with the following example.

Table 2 0.6-discernibility set for pairs of objects given in Table 1.

U	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}
x_1		$a_1a_2a_3a_4$	$a_1a_2a_3a_4$	$a_1a_2a_3a_4$	$a_1a_2a_3a_4$	$a_1a_2a_3a_4$	a_1a_4	a_4	$a_1a_2a_3a_4$	$a_1a_2a_4$
x_2			$a_1a_2a_3a_4$	$a_1a_3a_4$	a_1a_3	a_1a_3	a_1		a_1	a_2
x_3				a_1a_4	a_1a_4	a_1	a_1a_4		a_1	a_2
x_4			a_2a_3		a_2a_3	a_1a_3				a_2
x_5			$a_2a_3a_4$	$a_2a_3a_4$	$a_1a_2a_3a_4$		$a_1a_3a_4$	a_4	a_1	a_2
x_6	a_2		$a_2a_3a_4$	$a_2a_3a_4$	$a_1a_2a_3a_4$	$a_2a_3a_4$		a_4	a_2	a_2a_4
x_7	a_2a_3		$a_1a_2a_3a_4$	$a_1a_2a_3a_4$	$a_1a_2a_3a_4$	$a_1a_2a_3a_4$	$a_1a_2a_3a_4$		$a_1a_2a_3$	$a_2a_3a_4$
x_8	$a_1a_2a_3a_4$		$a_1a_2a_4$	$a_1a_2a_3a_4$	$a_1a_2a_3a_4$	$a_1a_2a_4$	$a_1a_2a_3a_4$		$a_1a_2a_3a_4$	$a_1a_2a_3a_4$
x_9	a_3a_4		$a_2a_3a_4$	$a_1a_2a_3a_4$	$a_1a_2a_3a_4$	$a_1a_2a_3a_4$	a_4			$a_2a_3a_4$
x_{10}	a_1a_3		$a_1a_2a_3a_4$	$a_1a_2a_3a_4$	$a_1a_2a_3a_4$	$a_1a_2a_3a_4$	a_1a_4		a_1a_4	

Example 1.3 (Continued from Example 1.2) For stochastic information system given in Table 1, following Example 1.2, the 0.6-discernibility sets for all pairs of objects are shown in Table 2. By the definition of discernibility matrix, one can

obtain the discernibility sets for all pairs of objects are shown in Table 2.

By Theorems 1 and 2, it is not difficult to obtain that $Red^{0.6}(S) = \{\{a_1, a_2, a_4\}\}$, i.e., $\{a_1, a_2, a_4\}$ is the only one 0.6-dominance reduct in Table 1.

4. ILLUSTRATIVE EXAMPLES

In this part, illustrative examples for the three cases are evaluated to express the potential implementation of the proposed approach.

Example 2

This is a modified example of selecting the most desirable strategy for an electricity retailer, which considered [35, 41]. In the example, there are nine objects (A_1, A_2, \dots, A_9) and four attribute: long-term profit (C_1), shortterm profit (C_2), market share (C_3) and green market share (C_4). Assume that the result of object A_i with respect to criterion C_j is a random variable with normal probability distribution $f_{ij}(x) = \frac{1}{\sigma_{ij}\sqrt{2\pi}} e^{-\frac{(x-\mu_{ij})^2}{2\sigma_{ij}^2}}$, which is given in Table 3. To determine the attribute reduct(s), computation processes and results using the proposed method are summarized as follows

Table 3 Means and standard deviations for consequences of objects with respect to attribute.

Objects	(μ_{ij}, σ_{ij})			
	C_1	C_2	C_3	C_4
A_1	(439,143)	(163,36)	(12.1, 0.5)	(9.3, 6.5)
A_2	(426,125)	(159,31)	(12.1, 0.5)	(14.8, 6.5)
A_3	(264,135)	(104,32)	(13.1, 0.5)	(9.3,6.5)
A_4	(444,125)	(163,31)	(12.1, 0.5)	(9.3, 6.5)
A_5	(605,115)	(220,31)	(11.0, 0.5)	(9.3, 6.5)
A_6	(449,126)	(166,32)	(12.1, 0.5)	(4.3, 6.5)
A_7	(449,107)	(164,27)	(12.1, 0.5)	(9.3, 6.5)
A_8	(457,126)	(165,32)	(12.1, 0.5)	(9.3, 6.5)
A_9	(453,107)	(163,27)	(12.1, 0.5)	(14.8, 6.5)

Firstly, using Eq. (7), four dominance degree matrices D_1, D_2, D_3 and D_4 are built as follows:

$$D_1 = \begin{bmatrix} 0.5 & 0.5273 & 0.8132 & 0.4895 & 0.1828 & 0.4791 & 0.4777 & 0.4624 & 0.4688 \\ 0.4727 & 0.5 & 0.8107 & 0.4594 & 0.1460 & 0.4484 & 0.4444 & 0.4307 & 0.4348 \\ 0.1868 & 0.1893 & 0.5 & 0.1640 & 0.0273 & 0.1582 & 0.1414 & 0.1480 & 0.1363 \\ 0.5105 & 0.5406 & 0.8360 & 0.5 & 0.1716 & 0.4888 & 0.4879 & 0.4708 & 0.4782 \\ 0.8172 & 0.8540 & 0.9728 & 0.8284 & 0.5 & 0.8198 & 0.8397 & 0.8072 & 0.8334 \\ 0.5209 & 0.5516 & 0.8418 & 0.5112 & 0.1802 & 0.5 & 0.5 & 0.4821 & 0.4903 \\ 0.5223 & 0.5556 & 0.8586 & 0.5121 & 0.1603 & 0.5 & 0.5 & 0.4807 & 0.4895 \\ 0.5376 & 0.5693 & 0.8520 & 0.5292 & 0.1928 & 0.5179 & 0.5193 & 0.5 & 0.5097 \\ 0.5312 & 0.5652 & 0.8637 & 0.5218 & 0.1666 & 0.5097 & 0.5105 & 0.4903 & 0.5 \end{bmatrix},$$

$$D_2 = \begin{bmatrix} 0.5 & 0.5336 & 0.8897 & 0.5 & 0.1151 & 0.4752 & 0.4911 & 0.4834 & 0.5 \\ 0.4665 & 0.5 & 0.8915 & 0.4637 & 0.0821 & 0.4376 & 0.4516 & 0.4464 & 0.4612 \\ 0.1103 & 0.1085 & 0.5 & 0.0927 & 0.0046 & 0.0853 & 0.0759 & 0.0888 & 0.0794 \\ 0.5 & 0.5363 & 0.9073 & 0.5 & 0.0968 & 0.4732 & 0.4903 & 0.4821 & 0.5 \\ 0.8849 & 0.9179 & 0.9954 & 0.9032 & 0.5 & 0.8873 & 0.9134 & 0.8915 & 0.9172 \\ 0.5248 & 0.5624 & 0.9147 & 0.5268 & 0.1128 & 0.5 & 0.5191 & 0.5088 & 0.5286 \\ 0.5089 & 0.5484 & 0.9241 & 0.5097 & 0.0866 & 0.4810 & 0.5 & 0.4905 & 0.5104 \\ 0.5166 & 0.5536 & 0.9112 & 0.5179 & 0.1085 & 0.4912 & 0.5095 & 0.5 & 0.5191 \\ 0.5 & 0.5388 & 0.9206 & 0.5 & 0.0828 & 0.4714 & 0.4896 & 0.4810 & 0.5 \end{bmatrix},$$

$$D_3 = \begin{bmatrix} 0.5000 & 0.5000 & 0.0787 & 0.5000 & 0.9401 & 0.5000 & 0.5000 & 0.5000 & 0.5000 \\ 0.5000 & 0.5000 & 0.0787 & 0.5000 & 0.9401 & 0.5000 & 0.5000 & 0.5000 & 0.5000 \\ 0.9214 & 0.9214 & 0.5000 & 0.9214 & 0.9985 & 0.9214 & 0.9214 & 0.9214 & 0.9214 \\ 0.5000 & 0.5000 & 0.0787 & 0.5000 & 0.9401 & 0.5000 & 0.5000 & 0.5000 & 0.5000 \\ 0.0599 & 0.0599 & 0.0015 & 0.0599 & 0.5000 & 0.0599 & 0.0599 & 0.0599 & 0.0599 \\ 0.5000 & 0.5000 & 0.0787 & 0.5000 & 0.9401 & 0.5000 & 0.5000 & 0.5000 & 0.5000 \\ 0.5000 & 0.5000 & 0.0787 & 0.5000 & 0.9401 & 0.5000 & 0.5000 & 0.5000 & 0.5000 \\ 0.5000 & 0.5000 & 0.0787 & 0.5000 & 0.9401 & 0.5000 & 0.5000 & 0.5000 & 0.5000 \\ 0.5000 & 0.5000 & 0.0787 & 0.5000 & 0.9401 & 0.5000 & 0.5000 & 0.5000 & 0.5000 \end{bmatrix},$$

$$D_4 = \begin{bmatrix} 0.5000 & 0.2748 & 0.5000 & 0.5000 & 0.5000 & 0.7068 & 0.5000 & 0.5000 & 0.2748 \\ 0.7252 & 0.5000 & 0.6265 & 0.5371 & 0.4734 & 0.5648 & 0.5303 & 0.5256 & 0.4740 \\ 0.5000 & 0.2748 & 0.5000 & 0.5000 & 0.5000 & 0.7068 & 0.5000 & 0.5000 & 0.2748 \\ 0.5000 & 0.2748 & 0.5000 & 0.5000 & 0.5000 & 0.7068 & 0.5000 & 0.5000 & 0.2748 \\ 0.5000 & 0.2748 & 0.5000 & 0.5000 & 0.5000 & 0.7068 & 0.5000 & 0.5000 & 0.2748 \\ 0.2932 & 0.1267 & 0.2932 & 0.2932 & 0.2932 & 0.5000 & 0.2932 & 0.2932 & 0.1267 \\ 0.5000 & 0.2748 & 0.5000 & 0.5000 & 0.5000 & 0.7068 & 0.5000 & 0.5000 & 0.2748 \\ 0.5000 & 0.2748 & 0.5000 & 0.5000 & 0.5000 & 0.7068 & 0.5000 & 0.5000 & 0.2748 \\ 0.7252 & 0.5000 & 0.6265 & 0.5371 & 0.4734 & 0.5648 & 0.5303 & 0.5256 & 0.4740 \end{bmatrix}.$$

Then, using Eq. (8), overall dominance degree matrix D can be obtained, i.e.,

$$D = \begin{bmatrix} 0.5000 & 0.4589 & 0.5704 & 0.4974 & 0.4345 & 0.5403 & 0.4922 & 0.4856 & 0.4359 \\ 0.5411 & 0.5000 & 0.6265 & 0.5371 & 0.4734 & 0.5648 & 0.5303 & 0.5256 & 0.4740 \\ 0.4296 & 0.3735 & 0.5000 & 0.4195 & 0.3826 & 0.4679 & 0.4097 & 0.4146 & 0.3530 \\ 0.5026 & 0.4629 & 0.5805 & 0.5000 & 0.4271 & 0.5422 & 0.4946 & 0.4882 & 0.4383 \\ 0.5655 & 0.5267 & 0.6174 & 0.5729 & 0.5000 & 0.6185 & 0.5783 & 0.5647 & 0.5213 \\ 0.4597 & 0.4352 & 0.5321 & 0.4478 & 0.3816 & 0.5000 & 0.4531 & 0.4460 & 0.4114 \\ 0.5078 & 0.4697 & 0.5904 & 0.5055 & 0.4218 & 0.5470 & 0.5000 & 0.4928 & 0.4437 \\ 0.5136 & 0.4744 & 0.5855 & 0.5118 & 0.4354 & 0.5540 & 0.5072 & 0.5000 & 0.4509 \\ 0.5641 & 0.5260 & 0.6471 & 0.5618 & 0.4787 & 0.5886 & 0.5563 & 0.5491 & 0.5000 \end{bmatrix}.$$

$D_1=0.52, D_2=0.47, D_3=0.60, D_4=0.51, D_5=0.43, D_6=0.56, D_7=0.51, D_8=0.50, D_9=0.45.$

In the following, we rank all objects according to the number of $D(i)$.

$$X_3 \succcurlyeq X_6 \succcurlyeq X_1 \succcurlyeq \begin{pmatrix} X_4 \\ X_7 \end{pmatrix} \succcurlyeq X_8 \succcurlyeq X_2 \succcurlyeq X_9 \succcurlyeq X_5.$$

If we consider the stochastic information system given in Table 3 and assume that a threshold of dominance degree 0.6, then we can obtain a Boolean matrix, which is corresponding to $D_{AT}^{0.6}$ such that:

$$D_{AT}^{0.6} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Take for instance x_2 , under the given dominance degree $\alpha = 0.6$, objects x_2 and x_3 are regarded as dominated by x_2 with respect to set of attributes AT. By the definition of discernibility matrix, one can obtain the discernibility sets for all pairs of objects are shown in Table 4.

Table 4 0.6-discriminability set for pairs of objects given in Table 3.

U									
	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9
x_1		$a_1a_2a_3$ a_4	a_3a a	$a_1a_2a_3$ a_4	a_1a_2 a_4	a_1a_2 a_3	$a_1a_2a_3$ a_4	$a_1a_2a_3$ a_4	$a_1a_2a_3$ a_4
x_2	$a_1a_2a_3$		a_3	$a_1a_2a_3$	a_1a_2 a_1a_2	a_3	$a_1a_2a_3$	$a_1a_2a_3$	a_4
x_3	$a_1a_2a_3$ $a_1a_2a_3$	$a_1a_2a_4$ $a_1a_2a_3$	a_3a	$a_1a_2a_4$	a_4 a_1a_2	a_1a_2 a_1a_2	$a_1a_2a_4$ $a_1a_2a_3$	$a_1a_2a_4$ $a_1a_2a_3$	$a_1a_2a_4$ $a_1a_2a_3$
x_4	a_4	a_4	a a_3a	a_4	a_4	a_3	a_4	a_4	a_4
x_5	a_3a_4 $a_1a_2a_3$	a_3a_4 $a_1a_2a_3$	a a_3a	a_3a_4 $a_1a_2a_3$	a_1a_2	a_3	a_3a_4 $a_1a_2a_3$	a_3a_4 $a_1a_2a_3$	a_3a_4 $a_1a_2a_3$
x_6	a_4 $a_1a_2a_3$	a_4 $a_1a_2a_3$	a a_3a	a_4	a_4 a_1a_2	a_4	a_4	a_4 $a_1a_2a_3$	a_4 $a_1a_2a_3$
x_7	a_4 $a_1a_2a_3$	a_4 $a_1a_2a_3$	a a_3a	$a_2a_3a_4$ $a_1a_2a_3$	a_4 a_1a_2	a_3 a_1a_2	a_4 $a_1a_2a_3$	a_4	a_4 $a_1a_2a_3$
x_8	a_4	a_4	a	a_4	a_4	a_3	a_4	a_4	
x_9	$a_1a_2a_3$	a_4	a_3	a_2a_3	a_1a_2	a_3	$a_1a_2a_3$	$a_1a_2a_3$	

By Theorems 1 and 2, it is not difficult to obtain that $Red^{0.6}(S) = \{\{a_1, a_3\}, \{a_2, a_3\}\}$, i.e., $\{a_1, a_3\}$ and $\{a_2, a_3\}$ are the 0.6-dominance reduct and $\{a_3\}$ is the core in Table 3.

Table 5. Discrete probability distributions and normal probability distributions for consequences of objects with respect to attribute.

Attribute	Objects	Scores										(μ_{ij}, σ_{ij})	
		1	2	3	4	5	6	7	8	9	10		
C_1	A_1	0	0.1	0.1	0.3	0.2	0.2	0.1	0	0	0		
	A_2	0	0	0.1	0.3	0.2	0.2	0.1	0.1	0	0		
	A_3	0.1	0	0.2	0.2	0.3	0.1	0.1	0	0	0		
	A_4	0	0	0	0.2	0.3	0.4	0.1	0	0	0		
C_2	A_1											(5.5,1.3)	
	A_2											(5,1.1)	
	A_3											(6,1.5)	
	A_4											(4.5,1)	
C_3	A_1	0	0.1	0.1	0.1	0.1	0.3	0.1	0.1	0	0		
	A_2												(4.5,1.2)
	A_3	0	0	0.2	0.3	0.2	0.2	0.1	0	0	0		
	A_4												(4.6,1.1)
C_4	A_1											(6,1.2)	
	A_2	0	0	0	0.1	0.2	0.3	0.2	0.2	0	0		
	A_3												(6.3,1)
	A_4	0	0	0	0	0.1	0.2	0.3	0.3	0.1	0		

Example 3

This is a modified example of (Lui et al.) choosing the most desirable vendor(s). There are four alternatives (A_1, A_2, A_3 and A_4) and four attribute: responsiveness to customer needs (C_1), price (C_2), on-time delivery (C_3) and product quality (C_4). Considerations of the alternatives

according to the attribute are shown in Table 5. To determine the attribute reduct(s), computation processes and results using the proposed method are summarized as follows

Firstly, using Equations (3), (4), (5), (6) and (7), four dominance degree matrices D_1, D_2, D_3 and D_4 are built as follows:

$$D_1 = \begin{bmatrix} 0.5 & 0.405 & 0.545 & 0.335 \\ 0.595 & 0.5 & 0.635 & 0.435 \\ 0.455 & 0.365 & 0.5 & 0.29 \\ 0.665 & 0.565 & 0.71 & 0.5 \end{bmatrix},$$

$$D_2 = \begin{bmatrix} 0.5 & 0.6155 & 0.4006 & 0.7290 \\ 0.3845 & 0.5 & 0.2954 & 0.6317 \\ 0.5994 & 0.7046 & 0.5 & 0.7973 \\ 0.2710 & 0.3683 & 0.2027 & 0.5 \end{bmatrix},$$

$$D_3 = \begin{bmatrix} 0.5 & 0.6125 & 0.575 & 0.5989 \\ 0.3875 & 0.5 & 0.4680 & 0.4755 \\ 0.425 & 0.5320 & 0.5 & 0.5090 \\ 0.4011 & 0.5245 & 0.4910 & 0.5 \end{bmatrix},$$

$$D_4 = \begin{bmatrix} 0.5 & 0.4548 & 0.42386 & 0.2554 \\ 0.5452 & 0.5 & 0.4778 & 0.31 \\ 0.5762 & 0.5222 & 0.5 & 0.3002 \\ 0.7446 & 0.69 & 0.6998 & 0.5 \end{bmatrix},$$

Then, using Eq. (8), overall dominance degree matrix D can be set up, i.e.

$$D = \begin{bmatrix} 0.5000 & 0.5220 & 0.4861 & 0.4796 \\ 0.4781 & 0.5000 & 0.4691 & 0.4631 \\ 0.5139 & 0.5310 & 0.5000 & 0.4741 \\ 0.5204 & 0.5370 & 0.5259 & 0.5000 \end{bmatrix}.$$

$$D_1=0.51, D_2=0.53, D_3=0.50, D_4=0.48.$$

In the following, we rank all objects according to the number of $D(i)$.

$$X_2 \succcurlyeq X_1 \succcurlyeq X_3 \succcurlyeq X_4.$$

If we consider the stochastic information system given in Table 5 and assume that a threshold of dominance degree 0.6, then we can obtain a Boolean matrix, which is corresponding to $D_{AT}^{0.6}$ such that:

$$D_{AT}^{0.6} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

Take for instance x_2 , under the given dominance degree $\alpha = 0.6$, objects x_2 is regarded as dominated by x_2 with respect to set of attributes AT .

By the definition of discernibility matrix, one can obtain the discernibility sets for all pairs of objects are shown in Table 6.

Table 6 0.6-discernibility set for pairs of objects given in Table 5.

U	x_1	x_2	x_3	x_4
x_1		a_1a_4	$a_1a_2a_3a_4$	$a_1a_3a_4$
x_2	$a_1a_2a_3a_4$		$a_2a_3a_4$	$a_1a_3a_4$
x_3	$a_1a_2a_3a_4$	$a_1a_3a_4$		$a_1a_3a_4$
x_4	a_2a_3	$a_1a_2a_3$	a_2a_3	

By Theorems 1 and 2, it is not difficult to obtain that $Red^{0.6}(S) = \{\{a_1, a_2\}, \{a_1, a_3\}, \{a_2, a_4\}, \{a_3, a_4\}\}$, i.e., $\{a_1, a_2\}, \{a_1, a_3\}, \{a_2, a_4\}$ and $\{a_3, a_4\}$ are the 0.6-dominance reduct in Table 5.

5. CONCLUSIONS

In this study, a general method is proposed to model the dominance relation in stochastic information systems. To check the dominance degree between two stochastic data, a threshold is utilized such α -dominance relation. This is useful in adapting information granulation according to varied levels.

In the framework, firstly, the dominance degree matrix of objects according to each attribute is established based on comparisons of probability distributions. The dominance degree provides that identifying process of stochastic dominance relations to compare objects. According to dominance relation, a rough set approach in stochastic information system is established as a substitution of the indiscernibility relation. Also we have discussed stochastic information systems in order to extract attribute reduct(s) based on the discernibility matrices. The proposed approach can be computed basic and easy procedures, beside the method provides to find much simpler attribute reduct(s) in a simple stochastic information system.

In terms of future study, the proposed model can be extended with multiple information forms such as crisp numbers, interval numbers, and stochastic information and so on.

REFERENCES

[1]. Pawlak, Z. (1982). Rough sets. *International Journal of Computer & Information Sciences*, 11(5), 341-356.
 [23]. Nowak, M. (2007). Aspiration level approach in stochastic MCDM problems. *European Journal of Operational Research*, 177(3), 1626-1640.
 [24]. Zaras, K., & Martel, J. M. (1994). Multiattribute analysis based on stochastic dominance (pp. 225-248). Springer Netherlands.
 [25]. Zaras, K. (2001). Rough approximation of a preference relation by a multi-attribute stochastic dominance for determinist and stochastic evaluation problems. *European Journal of Operational Research*, 130(2), 305-314.

[2]. Skowron, A., & Pawlak, Z. (2007). Rough sets: Some extensions. *Information Sciences*, 177(1), 28-40.
 [3]. Pawlak, Z. (1991). Rough sets: theoretical aspects of reasoning about data, system theory, Knowledge Engineering and Problem Solving, vol. 9.
 [4]. Hu, Q., Yu, D., Liu, J., & Wu, C. (2008). Neighborhood rough set based heterogeneous feature subset selection. *Information sciences*, 178(18), 3577-3594.
 [5]. Li, T., Ruan, D., Geert, W., Song, J., & Xu, Y. (2007). A rough sets based characteristic relation approach for dynamic attribute generalization in data mining. *Knowledge-Based Systems*, 20(5), 485-494.
 [6]. Qian, Y., Liang, J., Pedrycz, W., & Dang, C. (2010). Positive approximation: an accelerator for attribute reduction in rough set theory. *Artificial Intelligence*, 174(9), 597-618.
 [7]. Zhang, J., Li, T., Ruan, D., Gao, Z., & Zhao, C. (2012). A parallel method for computing rough set approximations. *Information Sciences*, 194, 209-223.
 [8]. Zhang, J., Li, T., Ruan, D., & Liu, D. (2012). Neighborhood rough sets for dynamic data mining. *International Journal of Intelligent Systems*, 27(4), 317-342.
 [9]. Guan, J. W., & Bell, D. A. (1998). Rough computational methods for information systems. *Artificial intelligence*, 105(1), 77-103.
 [10]. Qian, Y., Liang, J., & Dang, C. (2008). Converse approximation and rule extraction from decision tables in rough set theory. *Computers & Mathematics with Applications*, 55(8), 1754-1765.
 [11]. Greco, S., Matarazzo, B., & Slowinski, R. (1998, June). A new rough set approach to multicriteria and multiattribute classification. In *Rough sets and current trends in computing* (pp. 60-67). Springer Berlin Heidelberg.
 [12]. Greco, S., Matarazzo, B., & Slowinski, R. (2001). Rough sets theory for multicriteria decision analysis. *European journal of operational research*, 129(1), 1-47.
 [13]. Greco, S., Matarazzo, B., & Slowinski, R. (2002). Rough sets methodology for sorting problems in presence of multiple attributes and criteria. *European journal of operational research*, 138(2), 247-259.
 [14]. Greco, S., Matarazzo, B., Slowinski, R., & Stefanowski, J. (2000, October). An algorithm for induction of decision rules consistent with the dominance principle. In *Rough sets and current trends in computing* (pp. 304-313). Springer Berlin Heidelberg.
 [15]. Dembczyński, K., Pindur, R., & Susmaga, R. (2003). Generation of exhaustive set of rules within dominance-based rough set approach. *Electronic Notes in Theoretical Computer Science*, 82(4), 96-107.
 [16]. Dembczyński, K., Pindur, R., & Susmaga, R. (2003). Generation of exhaustive set of rules within dominance-based rough set approach. *Electronic Notes in Theoretical Computer Science*, 82(4), 96-107.
 [17]. Sai, Y., Yao, Y. Y., & Zhong, N. (2001). Data analysis and mining in ordered information tables. In *Data Mining, 2001. ICDM 2001, Proceedings IEEE International Conference on* (pp. 497-504). IEEE.
 [18]. Kotłowski, W., Dembczyński, K., Greco, S., & Slowiński, R. (2008). Stochastic dominance-based rough set model for ordinal classification. *Information Sciences*, 178(21), 4019-4037.
 [19]. J. Martel, K. Zaras, Dominance stochastique en analyse multicritère face au risque, Cahier No. 100. Laboratoire d'Analyse et Modélisation de Systèmes pour l'Aide à la Décision. Université Paris-Dauphine, Paris, 1990.
 [20]. J. Martel, S.H. Azondékon, K. Zaras, Preference relations in multicriterion analysis under risk, *Belgian Journal of Operations Research, Statistics and Computer Science* 31 (1992) 55–83.
 [21]. Nowak, M. (2004). Preference and veto thresholds in multicriteria analysis based on stochastic dominance. *European Journal of Operational Research*, 158(2), 339-350.
 [22]. Nowak, M. (2006). INSDECM—an interactive procedure for stochastic multicriteria decision problems. *European Journal of Operational Research*, 175(3), 1413-1430.
 [26]. Zaras, K. (2004). Rough approximation of a preference relation by a multi-attribute dominance for deterministic, stochastic and fuzzy decision problems. *European Journal of Operational Research*, 159(1), 196-206.
 [27]. Zawisza, M., & Trzpiot, G. (2002). *Multicriteria Analysis Based On Stochastic and Probabilistic Dominance in Measuring Quality of Life. In Multiple Objective and Goal Programming* (pp. 412-423). Physica-Verlag HD.

- [28]. Zhang, Y., Fan, Z. P., & Liu, Y. (2010). A method based on stochastic dominance degrees for stochastic multiple criteria decision making. *Computers & Industrial Engineering*, 58(4), 544-552.
- [29]. Lin, G., Qian, Y., & Li, J. (2012). NMGRS: Neighborhood-based multigranulation rough sets. *International Journal of Approximate Reasoning*, 53(7), 1080-1093.
- [30]. Qian, Y., Zhang, H., Sang, Y., & Liang, J. (2014). Multigranulation decision-theoretic rough sets. *International Journal of Approximate Reasoning*, 55(1), 225-237.
- [31]. Qian, Y., Li, S., Liang, J., Shi, Z., & Wang, F. (2014). Pessimistic rough set based decisions: a multigranulation fusion strategy. *Information Sciences*, 264, 196-210.
- [32]. Xu, W., Sun, W., Zhang, X., & Zhang, W. (2012). Multiple granulation rough set approach to ordered information systems. *International Journal of General Systems*, 41(5), 475-501.
- [33]. Yang, X., Qi, Y., Song, X., & Yang, J. (2013). Test cost sensitive multigranulation rough set: model and minimal cost selection. *Information Sciences*, 250, 184-199.
- [34]. Yang, X., Song, X., She, Y., & Yang, J. (2013). Hierarchy on multigranulation structures: a knowledge distance approach. *International Journal of General Systems*, 42(7), 754-773.
- [35]. Liu, Y., Fan, Z. P., & Zhang, Y. (2011). A method for stochastic multiple criteria decision making based on dominance degrees. *Information Sciences*, 181(19), 4139-4153.
- [36]. Graves, S. B., & Ringuet, J. L. (2009). Probabilistic dominance criteria for comparing uncertain alternatives: A tutorial. *Omega*, 37(2), 346-357.
- [37]. Zhang, W. X., & Qiu, G. F. (2005). *Uncertain decision making based on rough sets*. Publishing of Tsinghua University, Beijing.
- [38]. D. Bertsekas, J. Tsitsiklis, *Introduction to Probability*, Athena Scientific, 2002.
- [39]. W. Feller, *An Introduction to Probability Theory and Its applications*, Wiley, New York, 1971.
- [40]. Fouss, F., Achbany, Y., & Saerens, M. (2010). A probabilistic reputation model based on transaction ratings. *Information Sciences*, 180(11), 2095-2123.
- [41]. Lahdelma, R., Makkonen, S., & Salminen, P. (2006). Multivariate Gaussian criteria in SMAA. *European Journal of Operational Research*, 170(3), 957-970.
- [42]. Yager, R. R., Detyniecki, M., & Bouchon-Meunier, B. (2001). A context-dependent method for ordering fuzzy numbers using probabilities. *Information Sciences*, 138(1), 237-255.
- [43]. Qian, Y., Liang, J., & Dang, C. (2008). Interval ordered information systems. *Computers & Mathematics with Applications*, 56(8), 1994-2009.
- [44]. Yang, X., Qi, Y., Yu, D. J., Yu, H., & Yang, J. (2015). α -Dominance relation and rough sets in interval-valued information systems. *Information Sciences*, 294, 334-347.
- [45]. Shao, M. W., & Zhang, W. X. (2005). Dominance relation and rules in an incomplete ordered information system. *International journal of intelligent systems*, 20(1), 13-27.
- [46]. X.B. Yang, J.Y. Yang, C. Wu, D.J. Yu, Dominance-based rough set approach and knowledge reductions in incomplete ordered information system, *Inform. Sci.* 178 (2008) 1219–1234.

AOMDV Protokolünde Black Hole Ataklara Karşı Geliştirilmiş Güvenlik Uygulaması

Sinan TOKLU¹, Aziz AYDIN²

¹Bilgisayar Mühendisliği, Düzce Üniversitesi, Düzce, Türkiye

²Akademedy Yazılım, Kocaeli Üniversitesi Teknopark

sinantoklu@duzce.edu.tr, aziz.aydn@hotmail.com

(Geliş/Received:08.12.2016; Kabul/Accepted:03.03.2017)

DOI: 10.17671/gazibtd.309306

Özet- İletişimin büyük kısmını oluşturan kablosuz ağların kullanımı hızlı bir şekilde artmaktadır. Tasarsız ağlar sıklıkla hareketli düğümlerden oluşan alt yapısız kablosuz ağlardır. Tasarsız ağlarda düğümler hem diğer düğümlerle iletişim kurabilir hem de paketleri ileterek yönlendirici görevi üstlenirler. Bu ağlar arama kurtarma, ofis, kampüs, konferans salonu, üniversite ve şehir ağlarında kullanılmaktadırlar. Tasarsız ağlarda düğümlerin hareketli olması, bir altyapının mevcut olmaması, bant genişliğinin ve güç kapasitesinin sınırlı olması bu ağların en önemli sorunlarından. Düğümlerin hareketliliği topolojinin hızlı bir şekilde değişmesine ve kurulan yolların bozulmasına neden olmaktadır. Bu çalışmada, ağın güvenliğini bozacak black hole ataklar tasarlayıp, ağın güvenilirliğini artıracak güvenlik uygulaması geliştirilmiştir. Çalışmada tasarsız ağlarda kullanılan en güncel protokollerden birisi olan Ad hoc on-demand multipath distance vector (AOMDV) protokolü kullanılmıştır. Bu amaçla, güvenlik uygulamasında senaryolar üretilmiş olup, daha sonra da bu senaryolardan yararlanılarak AOMDV protokolünde güvenlik uygulamasının altyapısı oluşturulmuştur. Çalışmada benzetim aracı olarak Network Simulator (NS 2) programının 2.35 sürümü ve diğer ağ benzetim yazılımları için yardımcı programlar olan tracegraph202, APP-Tool-master grafik yazılımları kullanılmıştır.

Anahtar sözcükler- AOMDV, MANET, Black hole

Implemented Security Application Against Black Hole Attacks in AOMDV Protocol

Abstract- The use of wireless network forming the major part of the communication is increasing rapidly. Ad-hoc networks are typically composed of mobile nodes lower unstructured wireless networks. In ad-hoc network nodes can communicate with other nodes as they undertake the task of the router transmits both packages. These networks are search and rescue, office, campus, conference hall, university and city are used in the network. Be mobile nodes in ad-hoc networking, the absence of infrastructure, limited bandwidth and power capacity of the most important problem of this network. The topology of the nodes mobility leads to deterioration of rapid change and established ways. In this study, the design of black hole attacks to disrupt network security, is a security application that will increase the reliability of the network developed. Most of the current protocols used in networks tasarsiz study, which is one of AOMDV protocol was used. For this purpose, scenarios are being produced in a security application, and then utilizing the aomdv Protocol security in these scenarios the structure of the application has been established. In the study as a simulation tool the network Simulator (ns-2) Version 2.35 of the program, and other network utilities for the simulation software tracegraph202 app-tool-master software was used for graphics.

Keywords- AOMDV, MANET, Black hole

1. GİRİŞ (INTRODUCTION)

Tasarsız ağlar bir grup hareketli ya da hareketsiz düğümün bir araya gelerek oluşturduğu çok adımlı, önceden kurulmuş bir altyapıya sahip olmayan kablosuz ağlardır. Bu ağlardaki düğümler genelde hareketli olmakla beraber sabit düğümler de içerebilirler. Altyapılı ağlardaki gibi bir merkezi bir yönetim bulunmamaktadır. Düğümler hem yönlendirici görevini üstlenmekte hem de diğer düğümlerle iletişim kurmaktadır. Ağ altyapısız olduğundan düğümler istedikleri gibi hareket edebilmektedir. Tüm düğümler birbirlerinin kapsama alanında bulunamayacaklarından iletişim çok adımlı olarak yapılmaktadır. Tasarsız ağlar topoloji değişimlerine kolay uyum sağlayabilirler. Herhangi bir düğüm kurulan yoldan çıktığında durum fark edilir ve yeni bir yol kurma süreciyle iletişime kalınan yerden devam edilir. Bu durum gecikmeye sebep olsa da ağ hala iletişime imkan tanımaktadır. Tasarsız ağlardaki düğümlerin güç kaynakları, işlemci kabiliyetleri, saklama kapasiteleri ve bant genişlikleri kısıtlıdır. Güç kaynağının kısıtlı olması düğümlerin kapsama alanlarını sınırlandırmaktadır. Düğümlerin hareketli olması ağırlıklarına sınırlamalar getirmektedir. Bant genişliğinin sınırlı olması da tasarsız ağlarda gönderilecek kontrol mesajlarının sıklığına ve miktarına kısıtlamalar getirmektedir. Tasarsız ağlarda başarılı bir iletişim oluşabilmesi için bu kıt kaynakların etkin olarak kullanılabilmesi gerekmektedir [1].

Tüm uygulama alanlarının yönlendirme protokollerinden kendilerine has istekleri ve ihtiyaçları vardır. Duyarga ağ uygulamaları minimum enerji tüketimi isterken konferans uygulamaları gerçek zamanlı uygulamalar servis kalitesine önem vermektedirler. Tasarsız ağlarda kullanılan iletişim ortamının da -radyo haberleşmesi-kendine has özellikleri vardır. Örneğin, düğümler arasındaki bağlantılar tek yönlü olabilir. Bunun sebebi iki düğümün ileticilerinin güçlerinin farklılığı yüzünden sadece birinin diğerini duyabilmesi ya da ortamdaki gürültü olabilmektedir [2]. Çok adımlı iletişim yapılması hem güçte hem de iletim kapasitesinde yüksek kazançlara yol açmaktadır. Böylece düğümler paketleri çok daha az çıkış gücüyle çok adımda gönderebilmektedirler. Tasarsız ağların daha karmaşık uygulamalarda kullanılmasıyla, servis kalitesine (QoS) olan ihtiyaç artmaktadır. Bu ağların kullanılması ise güvenlik gereksinimlerini ortaya çıkarmıştır. Hareketli ad-hoc ağlarda (Mobile Ad-hoc Networks, MANET) gömülü bir güvenlik tasarımı yer almadığından ataklara karşı savunmasız yapıdadırlar. Dolayısıyla kablosuz kanal hem ağdaki kullanıcılara hem de ağda yer alan kötü niyetli kullanıcılara erişilebilir durumdadır. Güvenlik Ad hoc ağlarda ki en önemli konulardan biridir. Ad hoc ağlardaki en yaygın ataklar ağdan gönderilen paketlerin kötü niyetli düğümler tarafından yok edilmesi ve gelen paketlerde kötü niyetli düğümlerin değişiklik yaparak ağda karışıklığa yol açmasıyla ağın performansının düşürülmesinin hedeflenmesidir. MANET paylaşılan kablosuz ağ ortamı

vasıtasıyla hareketli düğümlerin birbiriyle iletişim isteğinde bulunduğu bir yapıya sahiptir [1, 3].

Bu çalışmada tasarsız ağların önemli bir sorunu olan güvenlik konusunda tasarsız ağlardaki güncel protokollerden biri olan AOMDV protokolü kullanılarak, ağda black hole saldırısı oluşturmak ve düğümlerin ağa katılımındaki hız parametreleri dikkate alınarak senaryolar üretip kötü niyetli düğümlere karşı ağda daha iyi performansın elde edilmesi için güvenlik uygulaması geliştirilmiştir. Bölüm 2’de AOMDV protokolü genel olarak anlatılmıştır. Bölüm 3’te ise gerçekleştirilen güvenlik uygulaması anlatılmıştır ve bölüm 4’te ise gerçekleştirilen benzetim ve elde edilen sonuçlar verilmiştir.

2. AOMDV PROTOKOLÜ (AOMDV PROTOCOL)

AOMDV protokolü Destination Sequenced Distance Vector (DSDV) protokolünden temel alınmıştır. AOMDV protokolünde on binlerce dinamik düğümden atlamalar yapılarak bir network sistemi oluşturulabilmektedir. AOMDV protokolünde temel konsept yön bulma süreçlerinde birçok yolun üretilmesi ve hesaplanmasıdır. Bağlantının kopması ve genellikle yanlış yönlendirmelerin gerçekleşmesi dinamik bir yapıya sahip olan AOMDV protokolünde avantaj sağlamaktadır. Yönlendirmede tek yol kullanan Ad Hoc On Demand Distance Vector (AODV) protokolü yönlendirmede herhangi bir yol kullanılmadığında yeni yol arayışına girmektedir. Bu durum her yol bulmada gecikmeye ve ağa yük getirisine neden olmaktadır. AOMDV protokolünde belirtilen verimsizliğin önüne geçmek için çoklu yol kullanılmaktadır[4, 5].

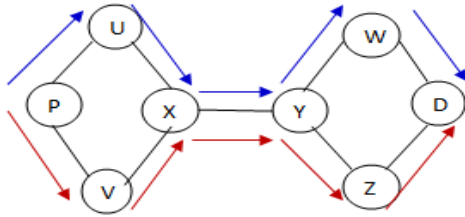
AOMDV protokolünde kaynak düğüm ağa istek paketi (Route Request,RREQ) yayar. Paketi alan ara düğüm yönlendirme tablosunu kontrol eder. Yönlendirme tablosunda hedef düğüme ulaşacak yol bilgisi mevcut ise kaynak düğüme geriye dönük cevap paketi (Route Reply,RREP) paketi gönderir. Hedef düğüme gidecek yol bilgisi yönlendirme tablosunda bulunmuyorsa komşu düğümlere RREQ paketi yaymaya devam edecektir. AODV Protokolünde paketi alan ara düğümde kopya RREQ paketleri oluştuğunda kopya RREQ paketlerden en geç gelen istek paketi göz ardı edilir. Hedef düğüme geç gelen RREQ paketi atılır. AOMDV protokolünde kopya RREQ paketleri atılmaz. Her kopya işlenir. AOMDV protokolünde hedef düğüme gelen paketler işlenmektedir. Gelen paketlerin çoklu olması, hedef düğümden gönderilen yayının diğer düğümlere birden çok yol sunması ve hangi yolların düğümlere ilan edilmesi ve düğümlerin hangi yolları kabul etmesi gerektiği sorusunu

AOMDV protokolünün loop freedom özelliği belirlenmektedir [6].

Döngüden kaçınmak için koşullar aşağıda listelenmiştir:

1. Farklı sıra numarası: Hedef düğüme gelen farklı sıra numarasına sahip paketlerden AODV protokolünde olduğu gibi en eski sıra numarasına sahip paket döngüden kaçınmak için atılır.
2. Aynı sıra numarası: Atlama sayısı daha kısa olan yol seçilir. Daha kısa yol bulunmazsa seçilen kısa yoldan başka seçenek bulunmamaktadır.

Hedef düğüme gidecek birden fazla ortak bağlantı bulunması ağda trafik oluşmasına ve tıkanıklığa neden olacaktır. Bundan dolayı ayrıık düğüm ve linkler gereklidir. AOMDV protokolünün path disjointness özelliği dikkate alınarak ağın performansı iyileştirilmektedir. Ayrıık bağlantı sağlama koşulunun gerçekleşmesi sonraki atlama ve son atlamanın birbirinden farklı olmasına bağlıdır. Şekil 2.1'de P düğümünden D düğüme giden yolda X düğümü belirtilen koşulu sağlamamaktadır. Bu durumda P-U-X-Y-W-D veya P-V-X-Y-Z-D yollarından biri kullanılacaktır [4, 7].



Şekil 2.1. Ayrıık bağlantı fikri. (Discrete connection idea)

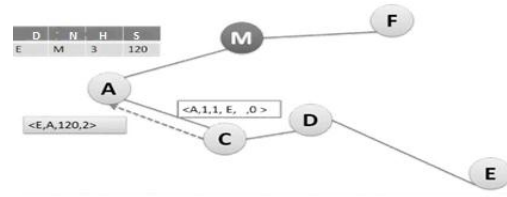
AOMDV protokolünde çoklu yol sayesinde RREP paketleri tüm düğümlere gönderilmektedir. Ad-hoc ağlarda bağlantı hataları hareketlilik, tıkanıklık, paket çakışmaları, düğüm ya da hop hataları vb. görülebilmektedir. Gönderilen paket yol boyunca bozuk bağlantıya rastlandığında hedef düğümden kaynak düğüme doğru RRER paketi yayılır. Yeni bir yol düğümlerin yönlendirme tablolarından tespit edilerek paket akışına devam edilir. Tüm bağlantı yolları bozulduğunda yeni bir yol kurmak gerekecektir. AODV protokolünde bozuk bağlantıya rastlanıldığında ağın kaynaklarını tüketerek yeni bir yol kurma arayışına gidilmektedir. AOMDV çoklu yol sayesinde bulunduğu ağın kaynaklarını daha verimli kullanabilmektedir [8].

AOMDV protokolünde yapılan çalışmalardan bir tanesi öncelikli AOMDV(P-AOMDV) olarak ta adlandırılan yeni bir yönlendirme protokolüdür. Bu protokol otobüs ulaşımı esnasında mobil altyapı kullanılarak beraber işlem yapabilme yeteneği kazandırmıştır. Bu sayede bağlantılabirlik kayıpların etkisi azaltılmış ve ağın performansı geliştirilmiştir. P-AOMDV kaynak ve hedef arasında bir çok yol bulmakta ve bu işlem sayesinde düğümlerin daha hızlı iletişimine destek vermektedir[9]. Diğer bir çalışmada ise çok yollu yönlendirme protokolü

geliştirilmiştir. Alıcı tabanlı AOMDV (RB-AOMDV) olarak adlandırılan bu protokol AOMDV protokolünün güvenilirlik avantajlarını almış ve tekrar kurulum keşif zamanını azaltmıştır. Alıcı düğüm belirli bir zaman aralığında herhangi bir veri paketi alamazsa yol keşfi işleminde olduğunu varsaymaktadır[10].

2.1. Black Hole Atak (Black Hole Attack)

Black hole atak Dos saldırı ataklarının türlerinden biridir. Uydurma yönlendirme bilgisi dağıtır ve üretir. Black hole saldırısında kötü niyetli bir düğüm uydurma yönlendirme bilgilerini gönderir. Kötü niyetli düğüm yönlendirme bilgisinin optimum olduğunu diğer düğümlere yayar. Saldırgan düğüm kaynak düğüme uydurma RREP paketi göndererek hedef düğüme giden daha iyi bir yol olduğunu iddia eder [11]. Hedef düğüme gönderilen RREP paketindeki hedef sıra numarası RREQ paketindeki sıra numarasından daha büyük veya eşit uydurma numarasıdır. Saldırgan kaynak düğümün kendisi üzerinden geçmesi gerektiği bilgisini yayarak ağdaki tüm trafiği yönetmektedir. Şekil 2.2'de RREP paketini alan kaynak düğüm örneği gösterilmiştir [12].



Şekil 2.2. RREP paketini alan kaynak düğüm örneği. (The source node instance that receives the RREP package)

Şekil 2.2'de gösterildiği gibi A kaynak düğümü E hedef düğüme veri göndermek istemektedir. Kötü niyetli düğümün (M) ağa etkisi tespit edilmiştir. Kaynak düğümde hedef düğüme gidecek yol bilgisi bulunmadığından, tüm komşu düğümlere RREQ paketi yayacaktır [13]. Gönderilecek RREQ paketinin yapısı aşağıdaki bilgilerden oluşmaktadır. Kaynak düğümün adresi, kaynak düğümün adresi, broadcast numarası, hedef düğümün adresi, hedef sıra numarası ve atlama sayısından oluşmaktadır. Kaynak düğüm RREQ paketini gönderdiğinde oluşacak RREQ paket içeriği <A,1,1,E, ,0> yapısında olacaktır. Paketi alan komşu düğümler (C,M) yönlendirme tablolarını kontrol edeceklerdir. C düğümü yönlendirme tablosunu kontrol ederek, E'ye giden bir yol olduğunu belirlerse geriye dönük A düğüme RREP paketi gönderecektir. Hedef düğüme giden yol bilgisi yönlendirme tablosunda mevcut değilse kaynak düğümden gelen bilgiyi yönlendirme tablosuna kaydedecektir. Komşu düğümlere istek paketi yaymaya devam edecektir.

Kötü niyetli düğüm paketi aldığı anda hedef düğüm bilgisinin kendisinde bulunduğu uydurma bilgisini kaynak düğümüne RREP paketini göndererek belirtir. Kötü niyetli düğüm hedef düğümüne kendisi üzerinden gidildiğinde daha az atlama sayısı ile ulaşabileceğini, RREP paketini kaynak düğümüne göndererek, kaynak düğümünden iletilen verinin kendisi üzerinden geçmesini istemektedir. Kaynak düğümüne gönderilecek RREP paket yapısı hedef düğümün adresi, gelecek düğümün adresi, atlama sayısı ve hedef düğümün sıra numarasından oluşmaktadır. RREP paket içeriği $\langle E, M, 120, 2 \rangle$ belirtildiği gibi olacaktır [14]. Kaynak düğüm RREP bilgisini aldıktan sonra veriyi M düğümüne gönderecektir. M düğümüne gelen paketler kaybolacaktır. Kötü niyetli düğüm amacına ulaşacaktır.

Bu alanda yapılan çalışmalardan bir tanesinde hiyerarşik kablosuz algılayıcı ağlar için farklı tipteki sinkhole düğümlerin belirlenmesi için yeni bir yaklaşım sunulmuştur. Bu yaklaşımda hiyerarşik kablosuz algılayıcı ağlar bir çok kümelere bölünmüştür ve her küme güçlü bir küme başına sahiptir. Bu küme başı görevinde olan düğümler kendi kümelerinde sinkhole atağı yapan düğümlerin belirlenmesi işlevini yerine getirmektedirler[15]. Diğer bir çalışmada ise Blackhole ve Grayhole atakların belirlenmesi için yeni bir mekanizma önerilmiştir. Bu mekanizma normal AODV yönlendirme protokolünün ve Watchdog mekanizmasının modifikasyonundan sonra çalışmaktadır. Bu mekanizmada düğüm veya herhangi bir yolun güvenilirlik seviyesinin tanımlanması için renk planını kullanmaktadır[16].

3. ÖNERİLEN YÖNTEM İLE AOMDV UYGULAMASI (AOMDV APPLICATION WITH THE PROPOSED METHOD)

Bu çalışmada kötü niyetli düğümlerin ağdaki saldırısının tüm ağı etkilemesinin azaltılması, dinamik yapıdaki düğümlerin belirli hızlarla ağa katılımının sağlanması, ağ kaynaklarının verimsiz kullanımını azaltan güvenlik tasarımı uygulaması amaçlanmış ve gerçekleştirilmiştir. Çalışmada ağdaki veri akışını bozabilecek düğümlerin ağ topolojisindeki karmaşıklık düzeyi incelenerek atakların etkisini azaltmaya dönük güvenlik uygulaması önerilmiştir.

AOMDV protokolünde black hole atağını gerçekleştirebilmek için ağa kötü niyetli düğüm eklenmiştir. Kötü niyetli düğümün ağa etkisi incelenmiştir. Kötü niyetli düğüm üzerinden yol kurulmuş, bu yol üzerinden geçen paketler atılmıştır. Kötü niyetli düğüm bunu yapmak için hedefe gidecek yolun kendisi üzerinden daha az atlama sayısı ile ve daha büyük sıra numarasıyla komşu düğümlere yanlış cevap olarak göndermektedir. Komşu düğümlerden gelen paketler kötü niyetli düğüm tarafından atılmaktadır. Diğer kablosuz yönlendirme protokolleri olan DSDV, AODV, DSR, HWMP vb. black hole saldırısına açıktır [17].

AOMDV protokolünde kötü niyetli düğümün davranışının ağın performansını düşürmesi ve ağdaki veri akışını bozması, dolayısıyla paket kaybının artmasını

engellemek için güvenlik uygulaması geliştirilmiştir. Düğümlerin ağa katılımında hareket özelliklerini kullanarak konumlanması ağın performansını artırmaktadır. Düğümlerin ağa belirli hızlarla katılımında kaynak düğümün komşu düğümlere istek paketi göndermeleri, paketleri alan komşu düğümlerin yönlendirme tablolarını güncellemesi ve hedef düğümüne gidecek yol bilgisi cevabı yer alan düğümlerin kaynak düğümüne cevap paketini göndermesiyle veri akışı başlamaktadır. Kötü niyetli düğümün ağdaki davranışı incelendiğinde düğümlerin ağa katılımında belirli hızlarla yer almaları ve hedef düğümüne gidecek sahte yönlendirme bilgisini kaynak düğümüne kolayca gönderebildiği görülmektedir [18]. Ağda düşük hızla konumlanan düğümler kötü niyetli düğümün paketleri kaybetmesine neden olmaktadır. Düğümlerin hız parametrelerinin artırılması ağda kurulacak yapının belirsizliği dolayısıyla kötü niyetli düğümlerin sahte yol bilgisi göndermelerini zorlaştırmaktadır.

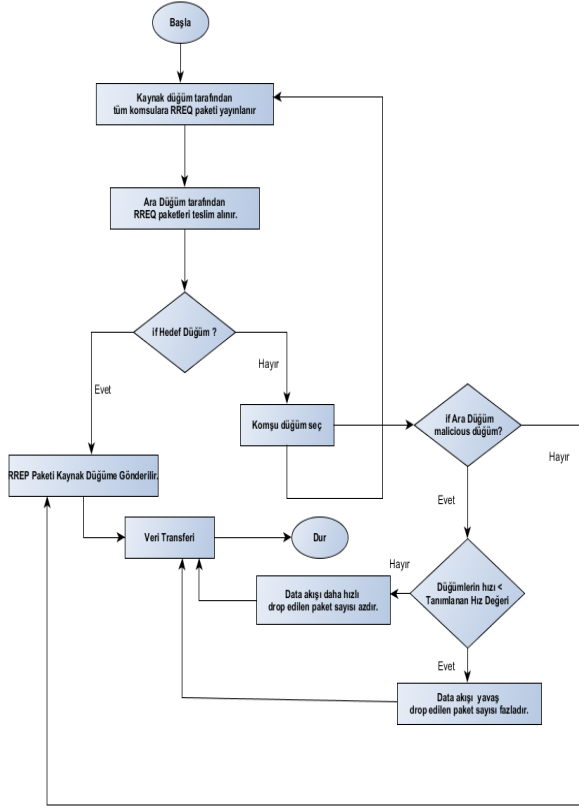
AOMDV protokolünde çoklu yol kullanıldığından hedef düğümüne giden yollardan birisi bozulduğunda hedef düğümüne ulaşan diğer yollar üzerinden veri akışı devam etmektedir. AOMDV protokolünün bu yapısı ağda yeni bir yol kurma arayışı için zaman harcamaması düğümlerin ağda konumlanmasının, kötü niyetli düğümün davranışının hız parametresi üzerinde değerlendirebilmekteyiz. Düğümlerin ağa katılım hız değerleri kontrol edilerek zaman aşımı maksimum hız değeri düğümlerin eşik değeri olarak tanımlanmaktadır. Eşik değere yaklaşan düğümlerin ağa katılma hızlarının eşik değere yaklaşması kötü niyetli düğümün ağdaki paket kaybına etkisini azaltmaktadır [19,20]. AOMDV protokolü güvenlik tasarımı yapısında tanımlanan hız değeri üzerinden ortaya çıkan akış diyagramı Şekil 3.1’de gösterilmiştir. Akış diyagramında tanımlanan hız değeri en yüksek atlama hız değeri olarak belirlenmiştir. Düğümlerin ağa katılım hız değerleri kontrol edilerek zaman aşımı maksimum hız değerini düğümlerin eşik değeri olarak tanımlamaktayız. Eşik değere yaklaşan düğümlerin ağa katılma hızlarının eşik değere yaklaşması kötü niyetli düğümün ağdaki paket kaybına etkisini azaltmaktadır. Eşik atlama hız değeri aşağıda tanımlanmaktadır.

$$l_b = E_n \text{ düşük atlama (0.2 m/s NS2.)}$$

$$u_b = E_n \text{ yüksek atlama (105 m/s NS2 [21].)}$$

Şekil 3.1’de gösterildiği gibi kaynak düğüm tüm komşu düğümlere paket yollamakta ve bu gönderilen paketler ara düğüm tarafından alınmaktadır. Eğer gönderilen paket hedefte ise ulaştığı bilgisi kaynak düğümüne gönderilmektedir. Gönderilen paket hedef düğümde değil ise komşu düğümlerden birisini seçer ve gelen paketi ona yönlendirir. Bu yeni ara düğüm kötü niyetli bir

düğüm ise ve düğümlerin hızı tanımlanan hız değerinden büyük olursa veri akışı hızlı olur ve atılan paket sayısı azalmaktadır. Fakat düğümlerin hızı tanımlanan hız değerinden küçük olursa bu durumda veri akışı yavaşlamakta ve atılan paket sayısı artmaktadır.



Şekil 3.1. Güvenlik uygulaması akış diyagramı.(Flow diagram of safety application)

3.1. Benzetim (Simulation)

Gerçekleştirilen uygulamanın benzetim aracı olarak Network Simulator (NS 2) programının 2.35 sürümü kullanılmıştır. NS kullanılması sebebi kullanıcılar çok sayıda fonksiyon sunması, kablosuz ağları desteklemesi ve açık kaynak kodlu olduğundan protokol eklemeye, mevcut protokolleri değiştirmeye imkân vermesidir. NS 2 kablolu ve kablosuz ağ araştırmalarında kullanılan nesneye dayalı bir ayrık olay benzetim aracıdır.

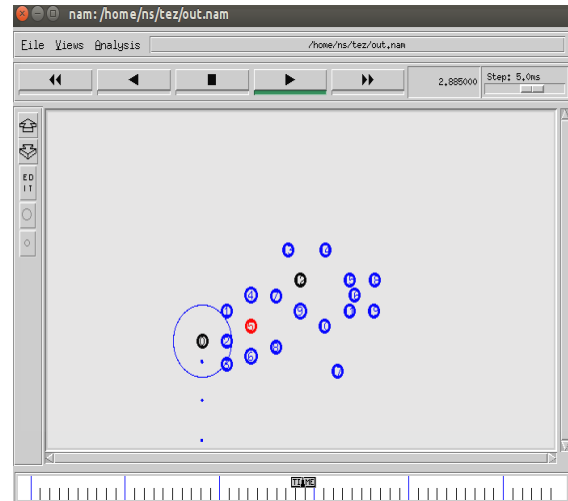
3.1.1. Benzetim Senaryoları (Simulation Scenarios)

Protokollerin performanslarının karşılaştırılması için 2 farklı senaryo, 6 farklı hareket modeline göre benzetim yapılmıştır. Tüm senaryolarda hareketli düğümler aynı atlama hızlarıyla alana rastgele dağıtılmıştır.

Tel dosyalarımız düğümlerin ağa katılımındaki hız değerleri dikkate alınarak oluşturulmuştur. Dosyalarımızda kötü niyetli düğümlerin ağa katılım hızı diğer düğümlerin hızlarıyla eş değer olarak belirlenmiştir. Düğümlerin ağa değişken hız tanımlarıyla katıldıklarında ağın performansını hangi ölçüde etkilediği benzetim aracıyla tespit edilmiştir.

Senaryolarımızda atlama hızları artan bir şekilde devam etmektedir. Senaryolarda tanımlanan hız değerleri maksimum hız değeri ve minimum hız değeri aralığında belirlenmektedir. Senaryo1 de tanımlanan atlama hızı değerini u1, senaryo 2 de tanımlanan hız değerini u2 belirtilmiştir. Güvenlik uygulaması senaryomuzda ağda yer alan beş numaralı düğüm kötü niyetli düğüm olarak belirlenmiştir. Kötü niyetli düğümün ağda görünmesi için düğüm farklı bir renkte tanımlanmıştır.

Düğümlerin ağdaki davranışları iki hareket senaryosu üzerinden incelenmiştir. Hareket modellerinde kaynak düğümden hedef düğüme giden yollar değişmektedir. İki farklı hareket senaryosunda hedef düğümler değişiklik göstermektedir. Ağa değişken atlama hızlarıyla katılan düğümlerin atlama hızlarında artış gerçekleştiğinde kötü niyetli düğümün ağdaki performansının etkisi benzetim aracıyla tespit edilmektedir. Şekil 3.2'de senaryolar için kullanılan topoloji gösterilmiştir.



Şekil 3.2. Senaryolar için kullanılan topoloji (Topology used for scenarios)

Şekil 3.2'de gösterilen topolojide ilk hareket modeli için 20 hareketli düğüm oluşturulmaktadır. Topoloji 700 X 700 m'lik bir alana kurulmuştur. Benzetim süresi 60 saniyedir. Hedef düğüm 12. düğüm olarak belirlenmiştir. İkinci hareket modeli 20 hareketli düğümden oluşmaktadır. Topoloji 800 X 800 m'lik bir alana kurulmuştur. Benzetim süresi 60 saniyedir. Hedef düğüm 17. düğüm olarak belirlenmiştir. Üçüncü hareket modeli 20 hareketli düğümden oluşmaktadır. Topoloji 1000*1000 m'lik bir alana kurulmuştur. Benzetim süresi 60 saniyedir. Hedef düğüm 19. düğüm olarak belirlenmiştir. Kaynak ve hedef düğüm aynı renklerde ve kötü niyetli düğüm farklı renkte tanımlanmıştır.

Hareketli düğümler “random waypoint” modeline göre hareket etmektedirler. NS 2 aracının düğüm hareket üretici “setdest” kullanılarak düğümlerin hareketlerini belirten hareket dosyaları oluşturulmuştur[8]. Hareket dosyası üretilirken ilgili senaryonun hareketli düğüm sayısı, maksimum hızı, benzetim süresi (60s), duraklama süresi ve topolojinin koordinatları verilmektedir. Benzetimde kullanılacak rastlantısal trafik bağlantıları NS 2 aracının trafik senaryo üretici “cbrgen” ile üretilmektedir. Her senaryo için kullanılan trafik parametreleri ve senaryolarda kullanılan genel parametreler Çizelge 3.1’de gösterilmiştir.

Çizelge 3.1. Benzetim parametreleri (Simulation parameters)

	Senaryo 1	Senaryo 2
Düğüm sayısı	20	20
Trafik tipi	CBR	CBR
Seed	1	1
Paket boyutu	512 Byte	512 Byte
Simülasyon süresi	60 s	60 s

3.1.2. Performans Kriterleri (Performance Criteria)

Ağa katılım ve hızları değişken olan düğümlerin katılım hızları ve varış noktalarına göre oluşturulan senaryoların kötü niyetli düğümün ağdaki performansının AOMDV protokolünde karşılaştırılması için aşağıdaki kriterler değerlendirilmiştir.

- Düşürülen paket miktarı: Kaynaklar tarafından üretilen veri paketi sayısının kötü niyetli düğüm tarafından düşürülmesidir. İletiminin verimi, düşürülen paket miktarı ve sonuç itibarıyla kullanılan yönlendirme protokolünün güvenilirliği hakkında bilgi sağlamaktadır.
- Ortalama gecikme: Gecikme bir veri paketinin varış tarafından alındığı zamandan paketin kaynak tarafından üretildiği zamanın çıkarılmasıyla elde edilmektedir. Gecikme, kuyrukta beklemekten, MAC seviyesindeki gecikmeden, iletim ve yayılım gecikmelerinden oluşmaktadır. Gecikme servis kalitesi için önemli bir parametre olduğundan kullanılan yönlendirme protokolünün güvenilirliği hakkında bilgi sağlamaktadır.

4. BENZETİM SONUÇLARI (SIMULATION RESULTS)

Bu bölümde önerilen 2 senaryonun ağa farklı katılım hızlarıyla yer almalarının kötü niyetli düğüm karşısındaki performansları benzetim sonuçlarına göre karşılaştırılmaktadır. Her senaryo için 2 farklı ağa katılım hızlarıyla oluşturulan ve kaynak düğümden hedef düğüme giden farklı yollar kullanılarak 6 farklı hareket modeline göre benzetim yapılmıştır.

4.1. Düşürülen Paket Miktarı (Dropped Package Amount)

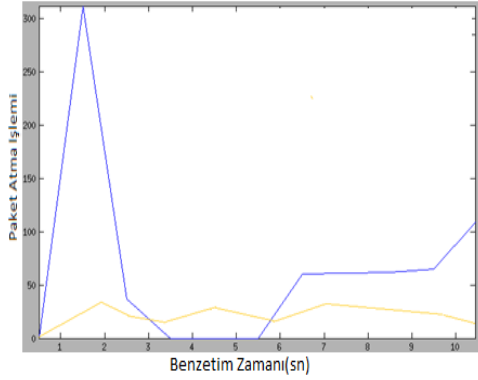
Şekil 4.1’de a’da gösterildiği gibi Senaryo 1-1 hareket modelinde 707 paket düşürülmüştür. Senaryo 2-1 hareket modelinde 204 paket düşürülmüştür. Kaynak düğüm 0. düğüm hedef düğüm 12.düğüm olarak belirlenmiştir. b’de ise düğümler Senaryo 1-2 hareket modelinde 692 paket düşürülmüştür. Senaryo 2-2 hareket modelinde 265 paket düşürülmüştür. Kaynak düğüm 0. düğüm hedef düğüm 17.düğüm olarak belirlenmiştir. c’de gösterilen grafikte ise Senaryo 1-3 hareket modelinde düğümlerden 769 paket düşürülmüştür. Senaryo 2-3 hareket modelinde 355 paket düşürülmüştür. Kaynak düğüm 0. düğüm hedef düğüm 19.düğüm olarak belirlenmiştir. Elde edilen sonuçlara bakıldığında düğümlerin farklı topolojilerde ağda değişken katılım hızlarıyla yer aldıklarında, ağa katılım hızının yüksek olması kötü niyetli düğümün paket atmasını zorlaştırmaktadır. Bu durum ağdaki paket kaybını azaltmakta ve ağın performansını artırmaktadır. Çizelge 3.2’de düşürülen paket miktarı tablo halinde gösterilmektedir.

Çizelge 3.2. Düşürülen paket miktarı (Dropped Package Amount)

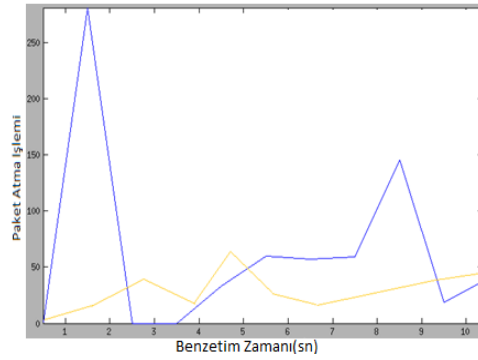
Senaryolar	Düşürülen paket miktarı	(Kaynak – Hedef) Düğüm
Senaryo 1-1	707	0-12
Senaryo 2-1	204	0-12
Senaryo 1-2	692	0-17
Senaryo 2-2	265	0-17
Senaryo 1-3	769	0-19
Senaryo 2-3	355	0-19

4.2. Ortalama Gecikme (Average Delay)

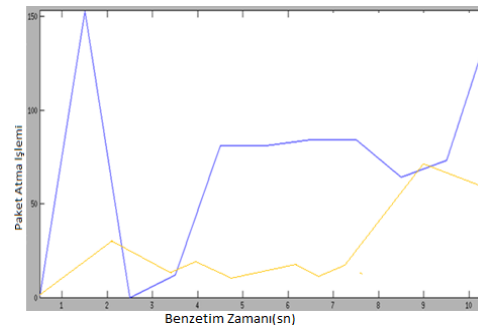
Şekil 4.2’de senaryo 1’de düğümlerin ağa ul katılım hızıyla katıldıklarında ortalama gecikme gösterilmektedir. Şekil 4.2 a’da gösterildiği gibi Senaryo 1-1 hareket modelinde ortalama gecikme 9873.62 ms’dir. Kaynak düğüm 0.düğüm hedef düğüm 12.düğüm olarak belirlenmiştir. b’de senaryo 1’de düğümlerin ağa belirlenen katılım hızıyla katıldıklarında ortalama gecikme gösterilmektedir. Kaynak düğüm 0. düğüm hedef düğüm 17.düğüm olarak belirlenmiştir. Senaryo 1-2 hareket modelinde ortalama gecikme 9275.3 ms’dir. c’de senaryo 1’de düğümlerin ağa belirlenen katılım hızıyla katıldıklarında ortalama gecikme gösterilmektedir.



a) Senaryo 1-1 ve Senaryo 2-1



b) Senaryo 1-2 ve Senaryo 2-2

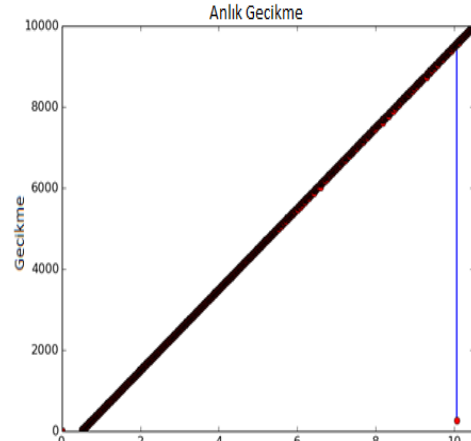


c) Senaryo 1-3 ve Senaryo 2-3

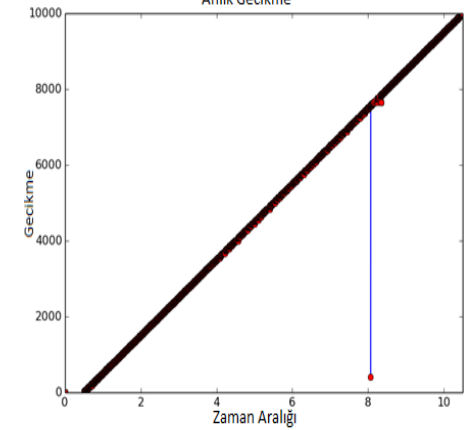
Şekil 4.1. İki Senaryonun Paket Düşürme Grafiği. (Package Dropped Chart of Two Scenarios.)

Çizelge 3.3. Ortalama gecikme (Average Delay)

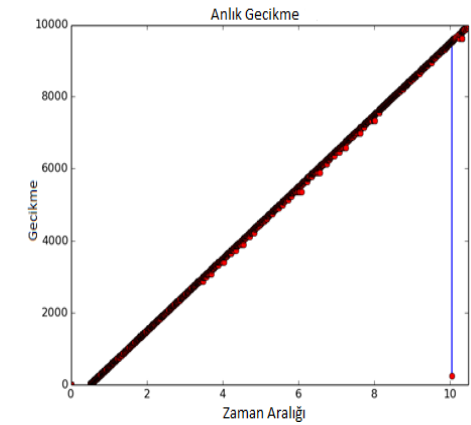
Senaryolar	Ortalama gecikme	(Kaynak – Hedef) Düğüm
Senaryo 1-1	9873.62 ms	0-12
Senaryo 2-1	6726.43 ms	0-12
Senaryo 1-2	9275.3 ms	0-17
Senaryo 2-2	7034.08 ms	0-17
Senaryo 1-3	9943.79 ms	0-19
Senaryo 2-3	7050.09 ms	0-19



a) Senaryo 1-1



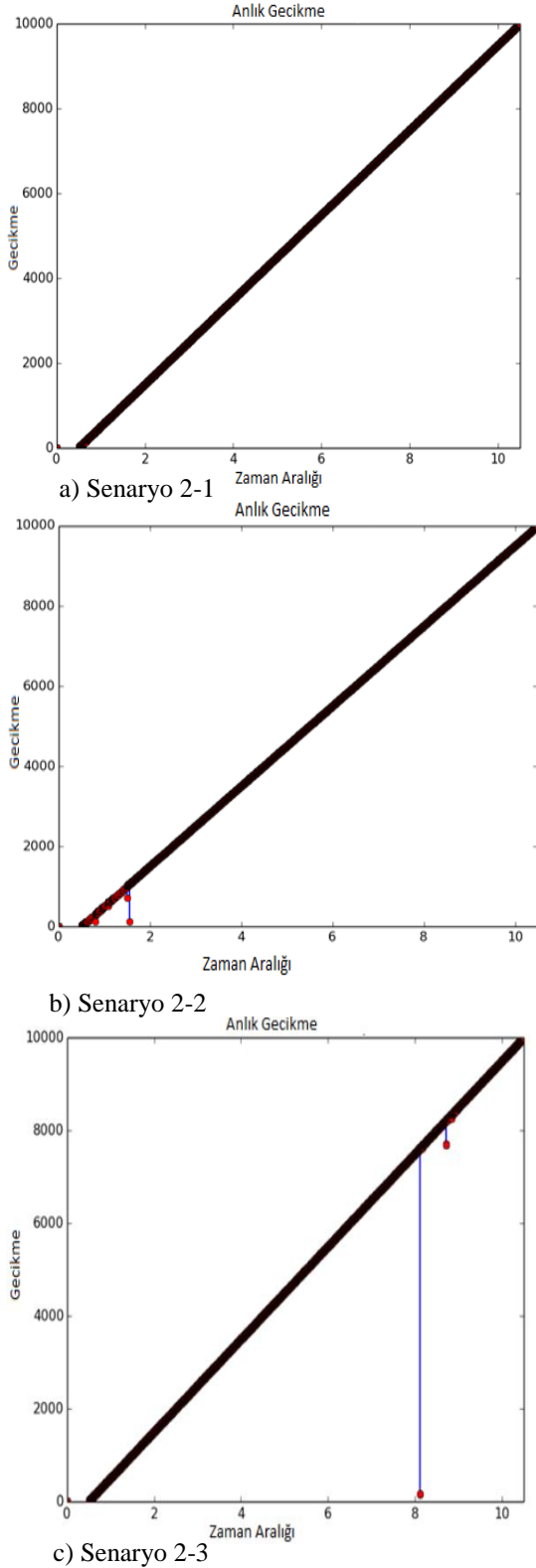
b) Senaryo 1-2



c) Senaryo 1-3

Şekil 4.2. Senaryo 1 ortalama gecikme süreleri. (Scenario 1 average delay times)

Kaynak düğüm 0. düğüm hedef düğüm 19.düğüm olarak belirlenmiştir. Senaryo 1-3 hareket modelinde ortalama gecikme 9943.79 ms'dir. Çizelge 3.3'de elde edilen ortalama gecikme tablo halinde sunulmaktadır. Şekil 4.3'de senaryo 2'de düğümlerin ağı u2 katılım hızıyla katıldıklarında ortalama gecikme gösterilmektedir.



Şekil 4.3. Senaryo 2 ortalama gecikme süreleri. (Scenario 2 average delay times.)

Şekil 4.3 a'da Senaryo 2-1 hareket modelinde ağa belirlenen katılım hızıyla katıldıklarında ortalama gecikme 6726.43 ms dir. Kaynak düğüm 0. düğüm hedef düğüm 12.düğüm olarak belirlenmiştir. b'de senaryo 2'de

düğümün ağa belirlenen katılım hızıyla katıldıklarında ortalama gecikme gösterilmektedir. Kaynak düğüm 0. düğüm hedef düğüm 17.düğüm olarak belirlenmiştir. Senaryo 2-2 hareket modelinde ortalama gecikme 7034.08 ms' dir. c'de ise senaryo 2'de düğümlerin ağa belirlenen katılım hızıyla katıldıklarında ortalama gecikme gösterilmektedir. Kaynak düğüm 0. düğüm hedef düğüm 19.düğüm olarak belirlenmiştir. Senaryo 2-3 hareket modelinde ortalama gecikme 7050.09 ms' dir.

6. SONUÇLAR (CONCLUSIONS)

Tasarsız ağlarda gömülü bir güvenlik tasarımı yer almadığından ataklara karşı savunmasız yapıdadır. Tasarsız ağların değişken topolojiye sahip olması, düğümlerin hareketli olması ve kablosuz ağ ortamından kaynaklanan olumsuzluklar birçok soruna neden olmaktadır. Dolayısıyla kablosuz kanal hem ağdaki kullanıcılara hem de ağda yer alan kötü niyetli kullanıcılara erişilebilir durumdadır. Tasarsız ağlarda düğümlerin hareketliliği ve topolojinin değişken olması yönlendirme işlemini bu ağların önemli problemlerinden biri kılmaktadır. Bu çalışmada güncel protokollerden AOMDV protokolü kullanılarak, Dos atak türlerinden Black hole saldırısı ağdaki güvenliği bozacak şekilde oluşturulmuştur. Düğümlerin ağa katılım hızları değerlendirilmiştir. Çalışmamızda maksimum hız değeri eşik değer olarak belirlenmektedir. Gerçekleştirdiğimiz benzetim senaryolarında düğümlerin ağa katılım hızlarında artış gerçekleştiğinde kötü niyetli düğümün daha az paket düşürdüğü ve ağdaki ortalama gecikmenin azaldığı elde edilen sonuçlarda gösterilmiştir.

Kötü niyetli düğümün davranışının ağın performansını düşürmesi ve ağdaki veri akışını bozması dolayısıyla paket kaybının artmasını engellemek için düğümlerin ağa katılım hızları artırılmıştır. Farklı topolojilerdeki ağ ortamlarında bulunan düğümlerin özellikle hareketli düğümlerin hedef noktalara veya hareket halinde iken hızlı davranmaları ağa yapılan saldırıların etkisini önemli ölçüde azaltmaktadır.

Tasarsız ağlarda topolojinin ağda değişkenlik göstermesi ve kaynakların sınırlı olması güvenlik uygulamalarının ortaya çıkmasını gerekli kılmaktadır. Topolojide oluşturulan Black hole saldırılarıyla ağın verimliliğinin benzetiminin yapılabildiği, ağa katılım hızlarında değişkenlik gerçekleştiğinde Black hole ataklara karşı düğümlerin direnç gösterdiği, düşürülen paket miktarları ve gecikmelerle tespit edilmiştir. Çalışmada tasarsız ağlarda saldırılara karşı dirençli olan güvenlik uygulamalarına ihtiyacın devam ettiği tespit edilmiştir.

KAYNAKLAR (REFERENCES)

- [1] Nevatia, Y., "Ad-Hoc Routing for USARSim", Networks and Distributed Systems Seminar, 2007.
- [2] Jani, P.V. "Security within AdHoc Networking," Position Paper, PAMPAS Workshop, 2002.
- [3] Gorantala, K. "Routing Protocols in Mobile Ad-hoc Networks", Master Thesis, UMEA University, 2006.
- [4] Gupta, P. ve Pandey, S. "Performance Analysis of AOMDV and AODV Routing Protocol in MANET", International Journal of Advanced Research in Computer Science and Software Engineering, vol.5(11), 2015.
- [5] Charles E. P. ve Elizabeth M. R. "Ad hoc on-demand distance vector routing", July 2003.
- [6] Tiwari, A. ve Verma, N. "A Novel Scheme for Intrusion Detection & Anticipation of Black Hole & Gray Hole Attacks In AODV Based MANET using ZED", International Journal Of Engineering And Computer Science ISSN:2319-7242 vol.3(5), 5744-5751, 2014.
- [7] Kurosawa, S. ve Nakayama, H. ve Kato, N. ve Jamalipour, A. ve Nemoto, Y. "Detecting Blackhole Attack on AODV-based Mobile Ad Hoc Networks by Dynamic Learning Method". International Journal of Network Security, vol.5(3), 338-346, 2007.
- [8] Zonglin, L. ve Guangming, H. ve Xingmiao, Y. "Spatial Correlation Detection of DDoS attack", International Conference on Communication, Circuits and System (ICCCAS 2009), 304-308, 2009.
- [9] Alves, J. ve Wille, ECG., "P-AOMDV: An improved routing protocol for V2V communication based on public transport backbones", Transactions on Emerging Telecommunications Technologies, vol:27(12), 1653-1663, 2016.
- [10] Al-Nahari, A., Mohamad, MM., "Receiver-Based Ad Hoc On Demand Multipath Routing Protocol for Mobile Ad Hoc Networks", PLOS ONE, vol:11(6), 2016.
- [11] Agrawal, S. ve Jain, S. ve Sharma, S. "A Survey of Routing Attacks and Security Measures in Mobile Ad-hoc Networks", Journal of Computing, vol.3(1), 2011.
- [12] Revathi, B. ve Geetha, D. "A Survey of Cooperative Black and Gray hole Attack in MANET", International Journal of C.S. And Management Research, vol 1(2), September 2012.
- [13] Gupta, H. ve Shrivastav, S. ve Sharma, S. "Detecting the DOS Attacks in AOMDV Using AOMDV-IDS Routing", International Conference on Computational Intelligence and Communication Networks, 2013.
- [14] Wang, K. ve Chen, J. ve Zhou, H. ve Qin, Y. "Content-Centric Networking: Effect of Content Caching on Mitigating DoS Attack", International Journal of Computer Science Issue, vol.9, 43-52, 2012.
- [15] Wazid, M., Das, AK., Kumari, S. ve Khan, MK., "Design of sinkhole node detection mechanism for hierarchical wireless sensor networks", Security and Communication Networks, vol:9(17), 45964614, 2016.
- [16] Khanna, N. ve Sharma, P., "Mitigating Blackhole and Grayhole Attack in MANET using Enhanced AODV with TLTB Mechanism", International Journal of Future Generation Communication and Networking, vol:9(8), 129-140, 2016.
- [17] Chadha, M. ve Joon R., ve Sandeep, "Simulation and Comparison of AODV, DSR and AOMDV Routing Protocols in MANETs", International Journal of Soft Computing and Engineering (IJSCE), vol 2(3), 2231-2307, 2012.
- [18] Bhalaji, N. ve Shanmugam, A. "A Trust Based Model to Mitigate Black Hole Attacks in DSR Based Manet", European Journal of Scientific Research ISSN 1450-216X, vol.50(1), 6-15, 2011.
- [19] Ahmed, M. ve Md. Hussain, A. "Performance of an IDS in an Adhoc Network under Black Hole and Gray Hole attacks" published in IEEE, 2014.
- [20] Chen H.L. ve Lee C.H., "Two Hops Backup Routing Protocol in Mobile Ad Hoc Networks", 11th International Conference on Parallel and Distributed Systems Workshops (ICPADS'05), 600-604, 2005.
- [21] Internet: <http://www.isi.edu/nsnam/ns/tutorial/>, 2016.

Bilgisayar Kullanıcılarına Yönelik Duygusal İfade Tespiti

Uğur AYVAZ, Hüseyin GÜRÜLER

Bilişim Sistemleri Mühendisliği, Muğla Sıtkı Koçman Üniversitesi, Muğla, Türkiye

ugur.ayvaz@mu.edu.tr, hguruler@mu.edu.tr

(Geliş/Received:12.12.2016; Kabul/Accepted:20.03.2017)

DOI: 10.17671/gazibtd.309307

Özet— Bilgisayar kullanımının yaygınlaştığı günümüzde, insan-bilgisayar etkileşimi ile ilgili yenilikçi çalışmalar hız kazanmıştır. Bu yeniliklerden bir tanesi de bilgisayar kullanıcısı bireylerin duygusal durumlarının, makine ile öğrenilmesidir. Ofis ortamlarında bilgisayarla çalışan bireylerin duygu durumlarının tespit edilebilmesi, özellikle bu kişilerin moral durumu ile iş performansı ilişkisi hakkında anlamlı bilgiler sunabilir. Bu fikirden hareketle, bilgisayar kullanıcısının yüz ifadelerine dayalı anlık duygu tespiti gerçekleştiren prototip bir sistem geliştirilmiştir. Geliştirilen bu sistem sırasıyla; yüz tespiti, yüz işaretçilerinin tespiti, yüz işaretçilerine dayalı özniteliklerden oluşan eğitim veri setinin oluşturulması ve kural-tabanlı sınıflandırıcı ile anlık duygusal durum tespitini gerçekleştirmektedir. Çalışmanın özgünlüğünü ifade eden özniteliklerin ayırt edici karakteristiğini anlamak amacıyla mevcut eğitim veri seti destek vektör makineleri ile durağan bir şekilde sınıflandırılmıştır. Sonuç olarak, sistemin başarımı 10-katlı çapraz doğrulama ile %96,1 olarak tespit edilmiştir.

Anahtar Kelimeler— İnsan-bilgisayar etkileşimi, duygusal ifade tespiti, veri madenciliği

The Detection of Emotional Expression towards Computer Users

Abstract— In these days, with the widespread use of computers, innovative studies in human-computer interaction have accelerated. One of these innovations is to learn one's emotional states with the help of a machine. Ascertaining the emotional states of a person working in the office environment can present meaningful information about the correlation between his/her morale and work performance. Based on this idea, an emotion detection system has been developed as a prototype. This system initially performs face and facial landmark detections. Training data set is prepared using these landmarks and then rule-based classifier detects instant emotional states. The training data set has been classified statically with support vector machines to understand the distinctive characteristics of the attributes that express the originality of this study. As a result, the success of the system was detected 96.1% with 10-fold cross-validation.

Keywords— Human-computer interface, detection of emotional states, data mining

1. GİRİŞ (INTRODUCTION)

İnsanlar arası yüz yüze etkileşim; sözlü, sözsüz ve sesli iletişim olmak üzere üç ana gruba ayrılmaktadır. Yapılan araştırmalara göre sözlü iletişim %7, sesli (ses tonu) iletişim %38, sözsüz iletişim (beden dili) ise %55 etkiye sahiptir [1, 2]. Bu sebeple; insan-makine etkileşiminde, insana ait ifadeleri anlama, algılama ve modellemede sözsüz iletişimden sıkça yararlanılmaktadır. Parmak izi tanıma [3], göz hareketi ve iris tanıma [4], el-yazısı tanıma [5], tuş vuruşu tanıma [6], hareket tanıma [7], el işareti tanıma [8], yüz tanıma [9] gibi sistemler bunlara örnek olarak verilebilir. Yüz tanıma, bireye dışarıdan herhangi bir müdahale olmadan kişisel kimliğini doğrulama ve modellemeye imkân sağlayan bir yöntemdir [10].

İnsan yüzü, sözsüz iletişim açısından vücudumuzda en fazla veriyi barındıran bölümdür. Yine insan yüzü, mimikler sayesinde insanlar arası etkileşimde olduğu gibi, insan-bilgisayar etkileşimi şeklinde de değerli bilgiler sunmaktadır. Örneğin, insan yüzlerine ait sayısal veri setleri bilgisayarlı sistemler ile işlenerek insan duygularının algılanması ve analizinde uzman sistemler olarak kullanılabilir. İnsan yüzü, sözsüz iletişim açısından vücudumuzda en fazla veriyi barındıran bölümdür. Yine insan yüzü, mimikler sayesinde insanlar arası etkileşimde olduğu gibi, insan-bilgisayar etkileşimi şeklinde de değerli bilgiler sunmaktadır. Örneğin, insan yüzlerine ait sayısal veri setleri bilgisayarlı sistemler ile işlenerek insan duygularının algılanması ve analizinde uzman sistemler olarak kullanılabilir.

Duygu üzerine yapılan çalışmalar kategorik ve boyutsal olmak üzere iki ana yaklaşımdan oluşmaktadır. Ekman ve diğer araştırmacıların savunduğu *kategorik* yaklaşımda duygu tanıma sadece yüze dayalı gerçekleştirilmekte ve duygular birbirinden ayrı bağımsız kategorilere ayrılmaktadır [11-14]. Bu yaklaşımda, altı duygu (mutluluk, şaşkınlık, üzüntü, öfke, iğrenme ve korku) evrensel kabul edilmektedir [11]. Yani bu duygular tüm

insanlar tarafından benzer yüz mimikleriyle ifade edilmektedir. Russell ve diğer araştırmacıların savunduğu *boyutsal* yaklaşımda ise; hoşnutsuzluk durumundaki sürekliliği ifade eden değerlik (duygusal ton) ve uyarılma seviyesi olmak üzere iki temel boyut (valence and arousal) bulunmaktadır [15-18]. İki boyutlu bu yaklaşım sadece yüze dayalı bir yaklaşım olmayıp, uyarılma durumu ve biyolojik sinyaller de duygu tanımında önem arz etmektedir [19]. Bu çalışmada ise Ekman ve diğerlerinin önerdikleri yüze dayalı duygu modelini kullanan bir sistem geliştirilmiştir.

Yüze dayalı duygu tanımının ilk adımı olan yüz tespiti için literatürde farklı araştırmacılar tarafından önerilmiş çeşitli yüz tespiti yöntemleri mevcuttur. Mikolajczyk ve diğerleri [20] vücut bölümlerinin tespit edilmesine yönelik geliştirdikleri çalışmada 7 adet vücut sensörü kullanarak yüzü, baş bölgesini, vücudun alt bölümlerini ve bacakları hem ön hem de yan profilden tespit etmişlerdir. Viola ve Jones [21] yüz tanıma sistemlerine farklı bir bakış kazandırmışlardır. Geliştirdikleri sınıflandırıcı algoritma, herhangi bir sensör gerektirmeden, video kareleri üzerinde gerçek zamanlı olarak yüz tespiti işlemine imkân vermektedir. Bu algoritma dijital resimleri ya da video karelerini 15fps (saniyedeki video kare sayısı) hızında analiz edip yüz ya da farklı bir nesnenin tespitini gerçekleştirebilmektedir. Algoritmanın çalışma hızını arttırabilmek mümkündür. Bu konuyla ilgili detaylı bilgi Viola-Jones algoritmasının anlatıldığı bölümde (2.1.) verilmiştir. Viola-Jones algoritmasının ön yüz tespitindeki başarısına karşın; yan profilden ve farklı çevresel şartlar altında (aydınlatma, oklüzyon, vb.) yüz tespiti gerçekleştirebilmek için Shen ve arkadaşları yeni bir yöntem geliştirmişlerdir. Bu yöntemde yüz tespitini ve yüz hizalama yöntemlerini bir arada kullanmışlardır [22]. Zhang ve arkadaşları [23], geçmişten günümüze kadar önerilen 12 farklı yüz tespiti yöntemlerinin performans değerlendirmesi üzerine çalışmalar gerçekleştirmişlerdir. Bu 12 yöntem; PASCAL, AFW, LFPW, FDDB gibi yüz resimleri içeren veri setleri üzerinde test edilmiştir. Hız ve doğruluk anlamında en iyi sonucu veren algoritma Chen ve arkadaşları [24] tarafından önerilen algoritma olmuştur. Chen ve arkadaşları, Viola-Jones algoritmasındaki cascade sınıflandırıcısı ile yüz işaretçileri hizalama tekniğini bir arada kullanarak bu alandaki en başarılı yüz tespit algoritmasını geliştirmişlerdir.

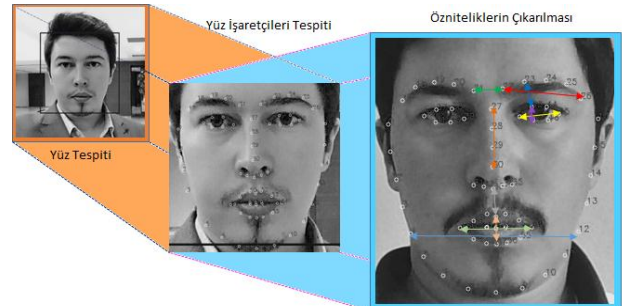
Yüze dayalı duygu tanıma ile ilgili olarak gerçekleştirilen çalışmalar incelendiğinde, Ekman ve diğerleri geliştirilen Yüz Eylem Kodlama Sistemi (YEKS) üzerinde yaptıkları çalışmalar ile insan yüzündeki kas yapısını ve hareketlerini görsel olarak incelemişler, bu kas hareketlerinin hangi duygusal ifadeleri temsil ettiğini tespit etmişlerdir [25, 26]. Essa ve Pentland klasik YEKS tabanlı statik resimlerle yapılan çalışmalara farklı bir yaklaşım getirmişlerdir. Bu yeni yaklaşımda video karelerinden elde ettikleri veri seti üzerinde yüzdeki kas hareketlerine göre, dinamik olarak, yüz ifadesi tespiti gerçekleştirmişlerdir [27]. Tsakanidou ve arkadaşları bir

yapısal ışık sensörü kullanarak gerçek zamanlı olarak kaydedilen iki ve üç boyutlu resimlerde otomatik yüz tanıma sistemi geliştirmişlerdir. Resimler üzerinde Active Shape Model tekniği ile 81 adet yüz işaretçisi takip işlemi gerçekleştirilmiştir. Bu işaretçiler kullanılarak yüze ait geometrik ve yüzey eğrilik ölçümleri hesaplanmış ve kural-tabanlı bir sınıflandırma yapılmıştır [28]. Zhang ve diğerleri [29] insansı robotların duygu tespiti yapabilmeleri amacı ile YEKS'e dayalı 54 yüz noktası elde etmişlerdir. Bir robota entegre edilen bu sistem, bulanık c-ortalama kümeleme (fuzzy c-means clustering) yöntemi ile duygu tespiti gerçekleştirmektedir. Chanthaphan ve diğerleri [30] derinlik sensöründen elde ettikleri yüze ait öznelikleri %91,17 doğruluk ile sınıflandırarak, duygu tespiti gerçekleştirmişlerdir.

Bu çalışmada, ofis ortamında çalışan bireylerin yüz ifadelerine dayalı duygusal durumları hakkında bilgi elde eden prototip bir bilişim sistemi geliştirilmiştir. Bu sistem, kategorik duygu tanıma baz alınarak, Viola-Jones'un yüz tespit algoritması üzerine inşa edilmiştir. Bu sistemi geliştirmekte yatan ana fikir; büyük işletmelerin ofislerinde çalışan personelin duygusal durumlarını nesnel bir şekilde anlamak, moral ve motivasyon durumlarını, yazılım destekli analiz edebilmek, iş performansı ve iş memnuniyeti açısından yöneticilere anlamlı bilgi sunabilecek bir alt yapıyı oluşturmaktır.

2. MATERYAL VE METOT (MATERIALS AND METHODS)

Bu çalışmada dizüstü bilgisayara birleşik video kamera ile 30 fps hızında gerçek zamanlı olarak yüz tanıma, yüz işaretçilerinin tespiti, öznelik çıkarımı ve duygusal ifade tespiti gerçekleştirilmiştir. Şekil 1'de gösterilen bu işlemlerin detayları ileriki bölümlerde tanıtılmaktadır. Çalışma; i5 2.90GHz işlemci, 8 GB RAM hafıza ve Nvidia GeForce GTX 860M grafik kartına sahip bir dizüstü bilgisayarda gerçekleştirilmiştir. Yazılım geliştirme aşamasında Python yazılım dili ve kütüphaneleri (OpenCV, Dlib, Scikit-learn, NumPy ve Orange) kullanılmıştır.



Şekil 1. Gerçek zamanlı yüz ifadeleri tanıma sisteminin ara basamakları

(Intermediate steps of real-time facial expression recognition system)

2.1. Yüz Tespiti (Face Detection)

Resim veya video gibi herhangi bir dijital ortamda yer alan insanlara ait yüzlerin, görüntü işleme teknikleri uygulanarak tespit edilmesi işlemidir.

Yüz tespit işleminde bütünsel ve analitik yaklaşımlar olmak üzere iki temel yaklaşım mevcuttur. Bütünsel yaklaşım yüzün bir bütün olarak ele alındığı yaklaşımdır. Huang ve Huang'ın [31] kullandıkları Nokta Dağılım Modeli (PDM) ile Pantic and Rothkrantz'ın [32] önerdikleri, resimlerdeki yüzü hem ön hem de yan profilden işleyen sistem bu yaklaşıma örnek olarak verilebilir. Analitik yaklaşım ise sadece ön yüze ait karakteristik yüz elemanlarının birlikte oluşumunu incelemektedir [33]. Kobayashi ve Hara'nın iris lokalizasyonu yöntemi [34] ile Rowley ve arkadaşlarının sinir ağları tabanlı yöntemine [35] ek olarak [21] ve [27] bu yaklaşıma örnek olarak verilebilir. Çalışmada Viola ve Jones [21] tarafından geliştirilen yüz tespit algoritması tercih edilmiştir. Bu algoritma; integral imaj, AdaBoost öğrenme algoritması ve Haar-cascade sınıflandırıcısı olmak üzere üç temel bölümden oluşmaktadır.

Haar-cascade sınıflandırıcı algoritması; dijital bir resim ya da video karesi içerisinde bulunan belirli bir nesnenin tespit edilmesi amacıyla kullanılmaktadır. Algoritmanın verimliliğini arttırmak amacıyla ilk olarak integral imaj üretilmektedir. İntegral imaj $ii(x, y)$ ile gösterilmektedir. Burada (x, y) imajın konumlandığı noktayı ifade eder. İntegral imaj denklem 1'de verilmiştir.

$$ii(x, y) = \sum_{x' \leq x} \sum_{y' \leq y} i(x', y'), \quad (1)$$

Denklemdaki $i(x, y)$, girdi olarak verilen imajdır. İntegral imaj, (x, y) koordinatının üstündeki ve sol tarafına doğru olan piksel değerlerinin toplamını içermektedir.

Haar-cascade sınıflandırıcı algoritmasının karar verme mekanizması denklem 2 ve 3'te verilmiştir.

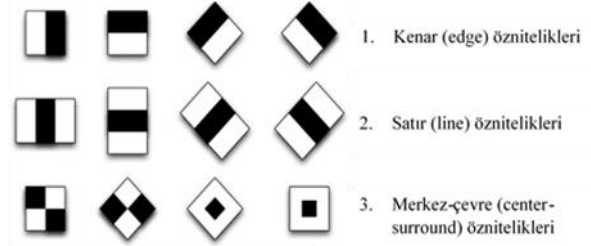
$$C_m = \begin{cases} 1, & \sum_{i=0}^{m-1} F_{m,i} > \theta_m \\ 0, & \text{otherwise} \end{cases} \quad (2)$$

$$F_{m,i} = \begin{cases} \alpha_{m,i}, & \text{if } f_{m,i} > t_{m,i} \\ \beta_{m,i}, & \text{otherwise} \end{cases} \quad (3)$$

Burada, $f_{m,i}$; 2-boyutlu integrallerin ağırlıklı toplamıdır ve aynı zamanda i . öznelik çıkarıcı için karar eşliğidir. $\alpha_{m,i}$ ve $\beta_{m,i}$; i . öznelik çıkarıcı ile bağımlı sabit değerlerdir. $\theta_{m,i}$; m . sınıflandırıcı için karar eşliğidir.

Yüz tespiti işleminde, Viola-Jones algoritmasını eğitmek amacıyla bir eğitim veri setine ihtiyaç vardır. Eğitim veri seti, hem içerisinde bulunması istenen nesnenin yer aldığı

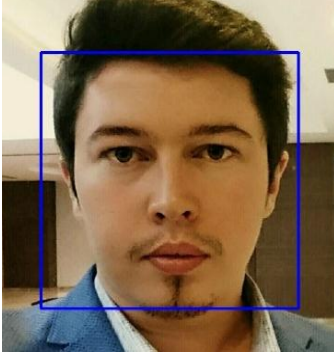
pozitif resimleri hem de yer almadığı negatif resimleri içermelidir. Pozitif resim içerisindeki nesneyi tespit etmek amacıyla Şekil 2'de görüldüğü gibi 24x24 piksel boyutlarındaki alt-pencereler tüm resim boyunca kaydırılarak nesne taraması gerçekleştirilir [21]. Bu pencereler, haar-benzeri öznelikler olarak bilinmektedir (Şekil 1). Tüm resim içerisinde bu küçük alt-pencerelerin kaydırılarak pozitif ve negatif bölgelerin tespit edilmesi çok zaman alıcıdır. Bu nedenle; Viola ve Jones, AdaBoost algoritmasını kullanarak son derece hızlı bir şekilde hem en iyi haar-öznelikleri seçmeyi hem de bu öznelikleri kullanan sınıflandırıcıları eğitmeyi başarmışlardır [36].



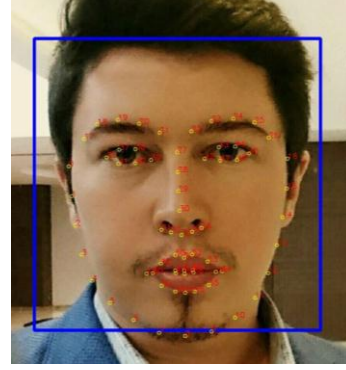
Şekil 2. Haar-benzeri öznelikler [37]
(Haar-like attributes)

Viola-Jones algoritması ilk geliştirildiğinde, çalışma hızı 15 fps'dir. Sonraki çalışmalar, grafik işleme birimi (GPU) ile merkezî işlem biriminin (CPU) birlikte kullanılması sayesinde Viola-Jones algoritmasının hızının artabileceğini göstermiştir. [38-40]. Gerçekleştirilen çalışmada Nvidia GeForce GTX 860M grafik kartına sahip bir bilgisayar kullanılmıştır. Bu grafik kartı 640 CUDA (Compute Unified Device Architecture-Birleşik Hesaplama Aygıt Mimarisi) çekirdeğine sahiptir. CUDA, hesaplamanın yoğun olduğu süreçte GPU'nun paralel hesaplama yeteneklerinden faydalanarak, tekrarlayan işlemlerin çalışma süresini kısaltmaktadır [39]. Open Source Computer Vision (OpenCV) kütüphanesi, Viola-Jones algoritmasının CUDA platformunda çalışabilmesini sağlamaktadır. Dolayısıyla Viola-Jones algoritması 30 fps hızında çalışabilmektedir. OpenCV'nin bu amaçla kullanıldığı çalışmalar literatürde yer almaktadır [40, 41].

Bu çalışmanın ilk adımında; Viola-Jones algoritması öncelikle statik resimlere uygulanmış ve yüz tespiti gerçekleştirilmiştir. Daha sonra, benzer işlemler gerçek zamanlı video karelerine uygulanarak video üzerinden yüz tespiti gerçekleştirilmiştir. Tespit edilen yüzler video boyunca takip edilmiştir. Gerçek zamanlı olarak yüz tespiti ve takibi işlemi 30 fps hızında gerçekleştirilmiş ve renkli bir çerçeve ile belirtilmiştir (Şekil 3).



Şekil 3. Gerçek zamanlı yüz tespiti
(Real-time face detection)



Şekil 4. Tespit edilen yüz üzerinde 68 adet yüz işaretçisinin gösterilmesi
(Displaying 68 facial landmarks on detected face)

2.2. Yüz İşaretçilerinin Tespiti (Detecting Facial Landmarks)

Yüz işaretçilerinin otomatik tespiti işlemi, yüz analizi çalışmaları için çok önemli bir yere sahiptir. Çünkü elde edilen yüz noktaları, yapılacak çalışmalar için zengin veriler sunmaktadır. Burada, tespit edilen noktaların doğruluğu ve hassasiyeti büyük önem arz etmektedir. Sagonas ve diğerleri mevcut; LFPW [43], HELEN [44], AFW [45] ve AFLW [46] veri setleri için kullanılan farklı yüz işaretçileri tespiti yöntemlerini ve yüz işaretçileri veri setlerini karşılaştırarak yeni bir yüz işaretçileri veri seti geliştirmişlerdir [42]. Önerdikleri yöntemde diğer veri setlerinden derledikleri 300-W [47] isimli bir veri seti ile çalışmışlar ve bir standart olarak kabul ettikleri 68 adet yüz işaretçisi veri setini elde etmişlerdir [42].

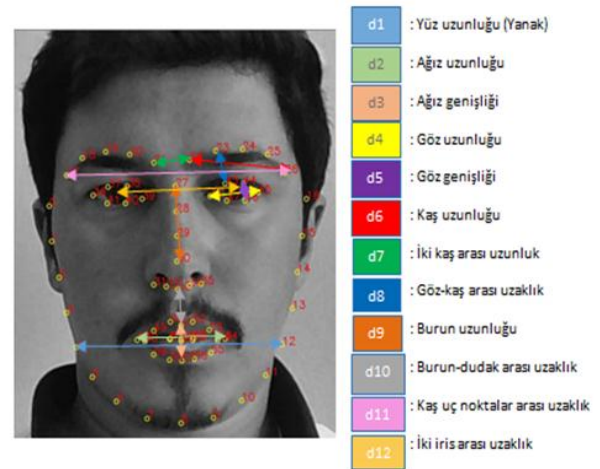
Gerçekleştirilen çalışmada yüz işaretçilerinin tespiti işleminde Sagonas ve arkadaşlarının önerdikleri ve bir standart kabul edilen veri seti tercih edilmiştir. C++ programlama dilinde geliştirilmiş platform-bağımsız bir yazılım kütüphanesi olan Dlib [48, 49], Sagonas ve diğerlerinin geliştirdikleri bu yüz işaretçisi veri setini içermektedir. Bu veri seti; insan yüz hatlarını ve yüz üzerindeki ağız, göz, kaş, burun gibi bölümlerin sınırlarını ifade eden 68 işaretçi bulundurmaktadır. Her bir işaretçi, yüz üzerinde belirli bir noktayı ifade etmektedir ve (x, y) koordinatlarına sahiptir. Bu işaretçiler; $P_1(x_1, y_1)$, $P_2(x_2, y_2)$, ..., $P_{68}(x_{68}, y_{68})$ şeklinde gösterilebilir.

Bu bölümde, gerçek zamanlı tespit edilen her bir yüz üzerinde, yine 30 fps hızında yüz işaretçileri tespit ve takip edilmiştir. Tespit edilen yüz işaretçileri ekranda 1'den 68'e kadar numaralandırılmış olarak Şekil 4'te gösterilmiştir. Bu yüz işaretçilerine ait koordinat değerleri kayıt altına alınarak incelendiğinde; kişinin cinsiyetine ve duygusal durum değişikliğine bağlı olarak koordinat değerlerinde belirgin değişiklikler olduğu görülmüştür. Buradan hareketle; tespit edilen yüz işaretçilerinden elde edilecek özniteliklerin, duygusal durumu ayırt etmede kullanılabileceği öngörülmüştür.

2.3. Öznitelik Çıkarımı (Attribute Extraction)

Özniteliklerin çıkarılması işlemi bu çalışmanın özgün olan yanısıdır, ayrıca duygu tanıma için en fazla önem arz eden süreçtir. Çünkü bu bölümde çıkarılan özniteliklerin doğruluğu ve ayırt edici özelliği, duygusal durumu sınıflandırma noktasında çok önemli bir role sahiptir. Bir önceki bölümde gerçek zamanlı olarak tespit ve takip edilen yüz işaretçileri arasındaki Öklid uzaklıkları belirlenerek 12 adet uzunluk elde edilmiştir. Bu uzunluk değerlerinin birbirine oranlanması ile 8 adet öznitelik çıkarılmıştır. Dokuzuncu öznitelik ise cinsiyet bilgisidir ve erkek için {0} ve kadın için {1} olarak tanımlanmıştır. Tablo 1'de tüm bu öznitelikler belirtilmiştir.

Yüz işaretçilerine dayalı çıkarılan öznitelikler; yüz bölümlerine ait 12 adet uzunluk değerinin (Şekil 5) birbirlerine oranıdır. Bu uzunluk değerlerinin seçilmesinde, farklı duygusal durumlara geçiş anlarında yüz işaretçilerinde meydana gelen gözle görülür konumsal değişiklikler etkili olmuştur. Elde edilen özniteliklerin duygusal durumu ayırt edici karakteristiği 2.6.1'de bahsedilen sınıflandırma işlemi ile doğrulanmıştır.



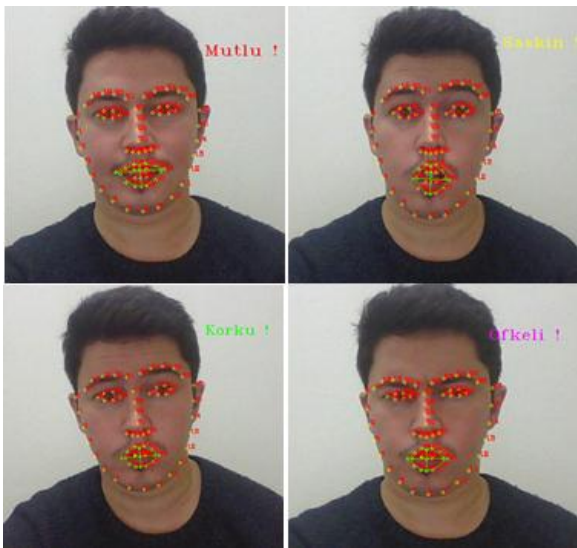
Şekil 5. Özniteliklerin belirlenmesinde kullanılan değerler
(Values used in determining the attributes)

Tablo 1. Yüz işaretçileri kullanılarak çıkarılan öznitelikler
(The attributes extracted using facial landmarks)

No	Öznitelik	Oran
1	Yüz genişliğinin ağız genişliğine oranı	d1/d2
2	Ağız genişliğinin ağız açıklığına oranı	d2/d3
3	Göz genişliğinin göz açıklığına oranı	d4/d5
4	Kaş uzunluğunun iki kaş arasındaki uzaklığa oranı	d6/d7
5	Kaş uzunluğunun göz-kaş arasındaki uzaklığa oranı	d6/d8
6	Burun uzunluğunun burun ile üst dudak arasındaki uzaklığa oranı	d9/d10
7	Kaşların en uç noktaları arası uzaklığın burun uzunluğuna oranı	d11/d9
8	İki iris arası uzaklığın ağız genişliğine oranı	d12/d2
9	Cinsiyet	{0,1}

2.4. Özniteliklere Dayalı Kural Oluşturma (Creating Rules For Instantaneous Detection Of Emotional Expressions)

Bu çalışmada özniteliklere dayalı kural oluşturulurken farklı bireylerin aynı duygusal durumdaki öznitelikleri kullanılmıştır. Altı farklı evrensel duygusal durum için bu işlem tekrarlanmıştır. Elde edilen özniteliklerin maksimum, minimum ve ortalama değerleri hesaplanarak altı farklı duygusal durum için eşik değerleri belirlenmiştir. Algoritma geliştirilirken kadınlardan ve erkeklerden elde edilen verilerin farklılık göstereceği göz önüne alınarak cinsiyete göre farklı eşik değerleri belirlenmiştir. Bu eşik değerlerine bağlı olarak yeni bir kural-tabanlı sınıflandırma algoritması geliştirilmiştir. “If-Else” yapısına dayalı bu algoritma ile gerçek zamanlı tespit edilen insan yüzünde altı farklı duygusal ifade yine anlık olarak tespit edilmiştir. Elde edilen sınıflandırma sonucu farklı renkler ile betimlenerek ekranda gösterilmiştir. Şekil 6’da çalışma kapsamında geliştirilen kural-tabanlı sınıflandırıcı ile tespit edilen altı farklı duygusal durumdan örnek olarak dördü gösterilmektedir.



Şekil 6. Duygusal ifadelerin anlık olarak tespitine örnekler

(Examples of instantaneous detection of emotional expressions)

2.5. Eğitim veri setinin oluşturulması (Creating training data set)

Bu çalışmanın özgün yanlarından bir tanesi de, eğitim amaçlı olarak hazır veri setlerini değil, özgün bir veri setini meydana getirmesidir. Bu bölümde, eğitim veri setinin hazırlanma süreci tanıtılmaktadır.

Gözetimli-denetimli öğrenmenin (Supervised Learning) bir gereği olarak hedef sütunu belli olan bir veri setine ihtiyaç vardır. Bu amaçla kamera karşısındaki 12 farklı kişiden (5 kadın, 7 erkek) 30 fps hızında 30 saniye süre ile yüz ifadelerine dayalı çıkarılan öznitelik verisi alınmıştır. Bu süre 5'er saniyelik alt bölümlere ayrılmış her bir bölümde kişilerin yalnızca belirli bir duygusal durumda kalmaları sağlanmıştır. Sonuç olarak, kayıt süresince her bir video karesi üzerinde yüz tespiti ve yüz işaretçileri tespiti işlemleri gerçekleştirilerek öznitelikler çıkarılmıştır. Şekil 7’de bir bölümü örnek olarak verilen veri seti; 9 adet öznitelik, 6 farklı sınıf ve toplamda 11452 adet veri örneğinden oluşmaktadır. Veri setinde kaydedilen 6 farklı sınıf, kişilerin kayıt anındaki gerçek duygusal durumlarıdır. Bu veri seti, kural-tabanlı sınıflandırma sonucu elde edilen tahmini duygusal durum bilgisini içermemektedir.

	ifade	d1/d2	d2/d3	d4/d5	d6/d7	d6/d8	d9/d10	d11/d9	d12/d2	cinsiyet
1	Mutlu	2.256	3.181	3.432	1.953	2.620	2.957	1.165	3.107	0
2	Mutlu	2.222	3.377	3.432	2.068	2.620	2.957	1.166	3.002	0
3	Saskin	2.466	1.381	2.722	2.762	2.530	2.209	1.503	2.983	1
4	Saskin	2.468	1.414	2.536	2.819	2.488	2.232	1.502	2.857	0
5	Uzgun	2.318	2.906	3.813	2.097	3.052	2.920	1.202	3.416	0
6	Ofke	2.189	6.596	3.761	3.861	2.578	4.606	1.235	2.807	0
7	Ofke	2.401	9.153	3.606	3.178	2.907	4.333	1.385	2.652	1
8	Tiksizme	2.507	2.479	4.617	1.969	2.910	3.170	1.254	3.016	0
9	Tiksizme	2.501	2.423	4.617	2.138	2.758	3.170	1.261	2.920	1
10	Korku	2.598	2.168	2.887	2.693	2.449	3.088	1.365	2.783	1
11	Korku	2.569	2.119	3.002	2.634	2.337	2.829	1.321	2.762	0
12	Korku	2.543	2.000	3.113	2.663	2.869	3.126	1.327	2.646	0

Şekil 7. Gerçek zamanlı duygusal durumlara ait öznitelikleri içeren veri setinden bir bölüm

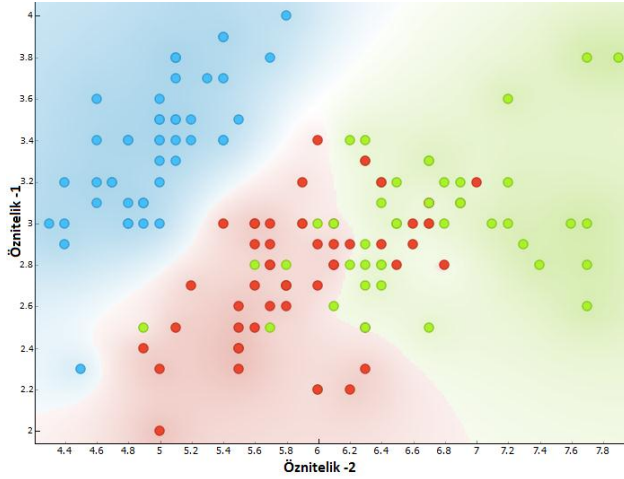
(A section from the data set containing attributes belonging to real-time emotional states)

2.6. Duygusal İfadelerin Statik Olarak Sınıflandırılması (Statically Classification Of Emotional Expressions)

Seçilen yüz işaretçilerine dayalı özniteliklerin ayırt etme karakteristiğini görebilmek amacıyla k-en yakın komşu (kNN) ve Destek Vektör Makineleri (DVM) kullanılarak 2.5 ‘te oluşturulan eğitim veri seti sınıflandırılmıştır. Bu iki algoritma doğrusal olarak sınıflandırılmayan ve çok sınıflı veri setlerinde başarılı sonuçlar vermektedir. Bu çalışmada, DVM algoritması ile kNN algoritmasına göre daha doğru sonuç elde edilmiştir. DVM ile ilgili detaylı bilgi 2.6.1’de verilmiştir.

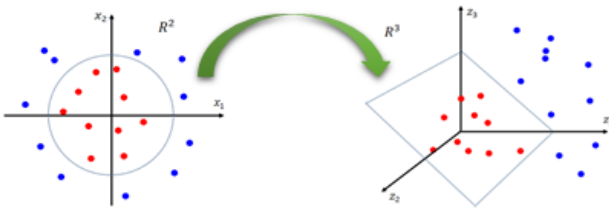
2.6.1. Sınıflandırmada destek vektör makinelerinin kullanımı (Usage of support vector machines in classification)

DVM, karmaşık ve büyük boyutlardaki veri setlerinin sınıflandırılmasında oldukça hızlı ve iyi performansa sahip bir denetimli öğrenme algoritmasıdır [50]. DVM algoritması iki veya çok sınıflı veri setlerinde, sınıfları birbirinden ayırmak için, bu sınıfların sınırlarını belirleyen bir düzlemin (hyperplane) veya düzlemlerin bulunması esasına dayanmaktadır [51]. DVM algoritması, verilerin hem doğrusal olarak ayrılabilirdiği hem de doğrusal olarak ayrılamadığı veri setlerinde kullanılabilir. Şekil 8’de üç-sınıflı herhangi bir veri setinin doğrusal olmayan DVM ile sınıflandırılması örnek olarak gösterilmiştir. Burada, veri setindeki herhangi iki özneliğe göre üç-sınıflı verilerin dağılımı doğrusal olmayan düzlemlerle üç farklı renk ile gösterilmiştir.



Şekil 8. Üç sınıflı bir veri seti örneğinin doğrusal olmayan DVM ile sınıflandırılması
(Classification of a sample of three-class data set with non-linear SVM)

Gerçekleştirilen çalışmada elde edilen veri seti altı-sınıflıdır ve doğrusal olarak ayrılamadığından sınıflandırma işlemi için doğrusal olmayan DVM algoritması tercih edilmiştir. Burada, veri setini doğrusal olmayan DVM ile sınıflandırabilmek için; veri seti iki-boyutlu uzaydan analitik uzaya taşınarak haritalama işlemi gerçekleştirilir [52]. Şekil 9’da R^2 , iki boyutlu girdi uzayını, R^3 , üç boyutlu özellik uzayı göstermektedir.



Şekil 9. Doğrusal olmayan verilerin iki-boyutlu uzaydan analitik uzaya dönüştürülmesi

(Transformation of non-linear data from two-dimensional space into analytic space)

$R^2 \rightarrow R^3$ dönüşümü sonucu elde edilen özellik uzayı; $(x_1, x_2) \rightarrow (z_1, z_2, z_3) := (x_1^2, \sqrt{2}x_1x_2, x_2^2)$ şeklinde ifade edilir. DVM sınıflandırıcısının karar fonksiyonu denklem 4’te verilmektedir.

$$\text{sgn}(\sum_{i=1}^m y_i \alpha_i k(x, x_i) + b) \quad (4)$$

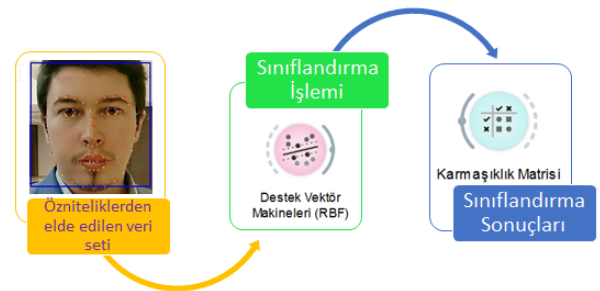
Burada, $k(x, x_i)$, DVM sınıflandırıcısının çekirdek fonksiyonunu ifade etmektedir.

DVM algoritmasında çekirdek fonksiyonu seçimi oldukça önemlidir. Aydoğan, gerçekleştirdiği çalışmada DVM çekirdek fonksiyonlarının performanslarını kıyaslamış ve farklı büyüklükteki farklı veri setlerinde en yüksek doğruluk oranlarını Radyal Bazlı Fonksiyon (RBF) ile elde etmiştir [53]. Bu çalışmada da DVM çekirdek fonksiyonu olarak RBF tercih edilmiştir. RBF aslında bir haritalama fonksiyonudur [54].

$$K_{RBF}(x, x') = \exp[-\gamma |x - x'|^2] \quad (5)$$

Denklem 5’te $x - x'$ veri setindeki iki farklı örneği temsil etmektedir. γ , bu iki örnek arasındaki benzerlik ölçüsünü vermektedir. $|x - x'|$ ifadesi ne kadar küçük olursa veri setindeki bu iki örnek birbirine o kadar yakın demektir. \exp , ifadesi ikili örnekler arası uzaklık bulma işleminin veri setindeki her bir örneğe uygulanarak gittikçe genişlemesi anlamına gelmektedir.

Statik sınıflandırma işlemi Orange (Ver.3.3.3) ile gerçekleştirilmiştir [55]. Sınıflandırma algoritmasının çalışma sistemi Şekil 10’da gösterilmiştir.



Şekil 10. Sınıflandırma algoritmasının çalışma sistemi
(The working system of classification algorithm)

DVM ile tüm veri seti, maksimum iterasyon sayısı 500 olacak şekilde sınıflandırıldığında %96,1 doğruluk elde edilmiştir. Bunun yanında, Tablo 2’de farklı örnek sayılarında, farklı sayıda iterasyonlar sonucunda elde edilen doğruluklar da verilmektedir. DVM algoritması kullanılarak geliştirilen sistemin başarımlarını değerlendirmesi 10-katlı çapraz doğrulama (10-fold cross validation) ile gerçekleştirilmiştir. Görüldüğü üzere; örnek sayısı arttıkça

daha fazla iterasyonla benzer doğruluk oranlarına ulaşabilmektedir.

Tablo 2. Sınıflandırma aşamalarında elde edilen sonuçlar
(Results obtained in classification stages)

Veri örnek sayısı	Maksimum iterasyon sayısı	Doğruluk (%)
6001	150	95.8
	250	96.6
	350	96.6
	500	96.6
8192	150	86.8
	250	95.0
	350	96.2
	500	96.2
9334	150	84.0
	250	94.2
	350	95.8
	500	96.0
11452	150	78.1
	250	89.9
	350	94.0
	500	96.1

Yüz tespiti üzerine yapılan yeni çalışmalarda [22, 45] farklı ışık koşullarında ve farklı yüz açılarındaki Viola-Jones algoritmasına göre daha yüksek doğrulukta yüz tespiti gerçekleştirildiği görülmektedir. Ayrıca bu yöntemlerin bir resim veya videoda yan profilden çekilmiş yüzleri tespit etmedeki başarısı Viola-Jones algoritmasına göre daha yüksektir. Ancak bu yöntemlerin hassasiyeti sağlayan algoritmalarındaki karmaşıklıkları nedeniyle Viola-Jones'a göre hızları ve gerçek zamanlı çalışma performansları düşüktür [24]. Bu çalışmanın yüz tespiti kısmında Viola-Jones algoritması diğer yüz tespiti algoritmalarına göre gerçek zamanlı çalışmalar için daha kullanışlı olduğundan tercih edilmiştir. Ayrıca, bu çalışmada önerilen sistem ofis ortamında bilgisayar ile çalışan bireylere yöneliktir. Birey bilgisayara bakarak çalıştığı süre boyunca yüzü ekrana dönük olacağından yüz tespitinde herhangi bir problem yaşanmamaktadır.

Çalışma kapsamında oluşturulan özgün eğitim veri setinde; 12 farklı bireyden (5 kadın, 7 erkek) 30'ar saniye süreyle 6 farklı duygusal durum için öznitelikler çıkarılmış ve 11452 adet veri örneğinden oluşan veri seti hazırlanmıştır. Bu veri seti; karmaşık ve büyük boyutlardaki veri setlerinin sınıflandırılmasında oldukça hızlı ve iyi performansa sahip bir denetimli öğrenme algoritması olan DVM algoritması ile sınıflandırılmıştır. DVM kullanan benzer çalışmaların [56-59] yüksek doğrulukta sonuçlar verdiği görülmüştür.

		Tahmin Edilen						
		Korku	Mutlu	Ofke	Saskin	Tikslenme	Uzgun	Σ
Gerçek	Korku	1807	15	21	20	2	15	1880
	Mutlu	2	1832	0	3	18	13	1868
	Ofke	12	0	1888	0	30	8	1938
	Saskin	30	0	20	1869	1	8	1928
	Tikslenme	8	13	80	5	1777	60	1943
	Uzgun	4	1	26	2	28	1834	1895
Σ		1863	1861	2035	1899	1856	1938	11452

Şekil 11. Sınıflandırma sonuçlarının karmaşıklık matrisi
(Confusion matrix of classification results)

Şekil 11'de verilen karmaşıklık matrisinde, DVM algoritması kullanılarak yapılan sınıflandırma işleminde doğru olarak sınıflandırılan toplam örnek sayısı 11007, yanlış sınıflandırılan toplam örnek sayısı ise 445 adettir.

4. TARTIŞMA (DISCUSSION)

Gerçekleştirilen çalışmada sırasıyla yüz tespiti, yüz işaretçileri tespiti, öznitelik çıkarımı ve duygusal durum sınıflandırma işlemleri uygulanmıştır. Yüz işaretçilerine dayalı özniteliklerin çıkarıldığı aşama özgünlük bakımından bu çalışma için en önemli aşamalardan biridir. Ayrıca elde edilen öznitelikler kullanılarak özgün bir eğitim veri seti de oluşturulmuştur.

Bu çalışmada gerçek zamanlı görüntü üzerinden sadece yüze dayalı duygu tanıma amaçlandığından Ekman'ın önerdiği ve duyguları kategorik olarak birbirinden ayıran model tercih edilmiştir.

3. SONUÇLAR (CONCLUSIONS)

Bu çalışmada ofis ortamında çalışan bireylerin yüz ifadelerine dayalı olarak duygusal durumları hakkında bilgi elde eden prototip bir bilişim sistemi geliştirilmiştir. Kişisel bilgisayara birleşik video kamera ile gerçek zamanlı olarak yüz tanıma, yüz işaretçilerinin tespiti, öznitelik çıkarımı ve duygusal ifade tespiti gerçekleştirilmiştir. Günümüzde, dizüstü bilgisayarlar için kameranın standart bir donanım haline gelmesi, bu sistemin kullanılabilirliğini artırmaktadır.

Bu çalışmanın yazılım tasarımında Python çatısı altında birleştirilen sayısal görüntü işleme ve makine öğrenmesi yöntemleri kullanılmıştır. Gerçek zamanlı duygu tespiti amaçlı geliştirilen kural-tabanlı sınıflandırıcı algoritma yüksek doğruluklarda çalışmaktadır. Aslında, belli bir duygu için standart bir kural oluşturmak oldukça zor bir süreçtir ve tekrarlayan "if-else" yapısına ihtiyaç duyulmaktadır. Elde edilen özniteliklerin ayırt edici özelliği ne kadar fazla ise bu özniteliklere dayalı kural oluşturmak o kadar kolaylaşmaktadır. Bu nedenle çıkarılan öznitelikler kural-tabanlı sistemin çalışmasını belirleyen en temel taşlardır. Çalışmanın kapsamı içerisinde, bu özniteliklerin ayırt edici değerini anlamak amacıyla eğitim veri seti doğrusal olmayan DVM algoritması ile sınıflandırılmıştır. DVM algoritmasının seçilme nedeni, özniteliklerin yüze ait noktalar arası Öklid uzaklığına bağlı çıkarılmış olmasıdır. kNN ve DVM algoritmaları bu tür konumsal değerlere sahip veri setlerinin sınıflandırılmasında oldukça başarılıdır. kNN ve DVM algoritmaları ile elde edilen doğruluk oranları

sırasıyla %95,7 ve %96,1 'dir. Burada doğrusal olmayan DVM algoritması ile elde edilen yüksek doğruluk değeri, çıkarılan özniteliklerin oldukça ayırt edici olduğu ve duygu sınıflandırmak için bu özniteliklere dayalı kural-tabanlı bir algoritmanın kullanılabilmesi yargılarını destekler niteliktedir. Bu araştırma neticesinde elde edilen sonuçlar, farklı öznitelikleri ve sınıflandırıcı yöntemleri araştırmaya kapı aralamaktadır.

Son olarak, büyük işletmelerde çalışan personelin duygusal durumlarının nesnel bir şekilde anlaşılması; işletme yöneticilerine anlamlı bilgiler sunulmasının yanısıra çalışanların duygusal durumlarının iş performansına etkisinin yazılım destekli analizine zemin oluşturmaktadır.

KAYNAKLAR (REFERENCES)

- [1] I. A. Essa, A. P. Pentland, "Coding, analysis, interpretation and recognition of facial expressions", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7), 757-763, 1997.
- [2] F. B. Mandal, S. Srivastava, "Real Time Facial Expression Recognition using a Novel Method", *The International Journal of Multimedia & Its Applications*, 4(2), 2012.
- [3] V. Conti et al., "Usability Analysis of a Novel Biometric Authentication Approach for Android-Based Mobile Devices", *Journal of Telecommunications and Information Technology*, 4, 34-43, 2014.
- [4] A.A. Mange et al., "Gaze and blinking base human machine interaction system", **IEEE International Conference on Computational Intelligence and Computing Research (ICIC)**, 2015.
- [5] C. O'Reilly et al., "Recent developments in the study of rapid human movements with the kinematic theory: Applications to handwriting and signature synthesis", *Pattern Recognition Letters*, 35(1), 224-235, 2014.
- [6] A. Salem et al., "Analysis of strong password using keystroke dynamics authentication in touch screen devices", **Cybersecurity and Cyberforensics Conference CCC 2016**, 15-21, 2016.
- [7] K. Sehairi et al., "Real-time implementation of human action recognition system based on motion analysis", *Studies in Computational Intelligence*, 143-164, 2016.
- [8] G. Castellano et al., "Recognising Human Emotions from Body Movement and Gesture Dynamics", **2nd international conference on Affective Computing and Intelligent Interaction**, 71-82, 2007.
- [9] S. Borra and N.K. Jayant, "Attendance Management System Using Hybrid Face Recognition Techniques", **Conference on Advances in Signal Processing (CASP2016)**, 2016.
- [10] M. Moon and J.P. Phillips, "Computational and performance aspects of PCA-based face recognition algorithm", *Perception*, 30(3), 303-321, 2001.
- [11] P. Ekman, D. Keltner, "Universal Facial Expression of Emotion: An Old Controversy and New Findings", *J Nonverbal Behav*, 21(1), 3-21, 1997.
- [12] P. Ekman, "**Basic Emotions**", Hand Book of Cognition and Emotion, Editör: T. Dalgleish ve M. J. Power, Wiley, New York, A.B.D., 45-60, 1999.
- [13] N. L. Ectoff, J. J. Magee, "Categorical perception of facial expressions.", *Cognition*, 44, 227-240, 1992.
- [14] D. Roberson, L. Damjanovic, M. Kikutani, "Show and tell: The role of language in categorizing facial expressions of emotion.", *Emotion Review*, 2, 255-260, 2010.
- [15] J. A. Russell, "A circumplex model of affect." *Journal of Personality and Social Psychology*, 39, 1161-1178, 1980.
- [16] J. A. Russell, "Core affect and the psychological construction of emotion.", *Psychological Review*, 110(1), 145-172, 2003.
- [17] M. Katsikitis, "The classification of facial expressions of emotion: A multidimensional-scaling approach.", *Perception*, 26, 613-626, 1997.
- [18] T. Takehara, N. Suzuki, "Differential processes of emotion space over time.", *North American Journal of Psychology*, 3, 217-228, 2001.
- [19] T. Fujimura, Y.T. Matsuda, "Categorical and dimensional perceptions in decoding emotional facial expressions", *Cognition and Emotion*, 26(4), 587-601, 2012.
- [20] K. Mikolajczyk, C. Schmid, A. Zisserman, "Human detection based on a probabilistic assembly of robust part detectors", **Proc. of ECCV**, 2004.
- [21] P. A. Viola, M. J. Jones, "Robust Real-Time Face Detection", *Int J Comput Vision*, 57(2), 137-154, 2004.
- [22] X. Shen, Z. Lin, J. Brandt, Y. Wu, Detecting and aligning faces by image retrieval, **2013 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)**, 3460-3467, 2013.
- [23] Z. Zhang et al., "A survey on face detection in the wild: Past, present and future", *Computer Vision and Image Understanding*, 138, 1-24, 2015.
- [24] D. Chen, S. Ren, Y. Wei, X. Cao, J. Sun, "Joint cascade face detection and alignment.", **European Conference on Computer Vision (ECCV) 2014**, 2014.
- [25] P. Ekman, W. V. Friesen, "Measuring facial movement", *Environmental psychology and nonverbal behavior*, 1(1), 56-75, 1976.
- [26] G. Donato, M. S. Bartlett, J. C. Hager, P. Ekman, T. J. Sejnowski, "Classifying facial actions", *IEEE Trans. Pattern Anal. Mach. Intell.*, 21(10), 974-989, 1999.
- [27] I. A. Essa, A. P. Pentland, "Facial Expression Recognition using a Dynamic Model and Motion Energy", **International Conference on Computer Vision '95**, Cambridge, MA, 360-367, 1995.
- [28] F. Tsalakanidou, S. Malassiotis, "Real-time 2D+ 3D facial action and expression recognition", *Pattern Recognition*, 43(5), 1763-1775, 2010.
- [29] L. Zhang et al., "Adaptive facial point detection and emotion recognition for a humanoid robot", *Computer Vision and Image Understanding*, 140, 93-114, 2015.
- [30] Chanthaphan et al., "Novel facial feature extraction technique for facial emotion recognition system by using depth sensor", *International Journal of Innovative Computing*, 12(6), 2067-2087, 2016.
- [31] C.L. Huang and Y.M. Huang, "Facial Expression Recognition Using Model-Based Feature Extraction and Action Parameters Classification." *J. Visual Comm. and Image Representation*, 8(3), 278-290, 1997.

- [32] M. Pantic and L. Rothkrantz, "Expert System for Automatic Analysis of Facial Expression", *Image and Vision Computing J.*, 18(11), 881-905, 2000.
- [33] P.S. Subramanyam, S.A. Fegade, "Face and Facial Expression Recognition - A Comparative Study", *International Journal of Computer Science and Mobile Computing*, 2(1), 2013.
- [34] H. Kobayashi and F. Hara, "Facial Interaction between Animated 3D Face Robot and Human Beings", *Proc. International Conf. on Systems, Man, Cybernetics*, 4, 3732-3737, 1997.
- [35] H. Rowley, S. Baluja, T. Kanade, "Neural Network-Based Face Detection", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20,(1), 23 – 38, 1998.
- [36] Y. Freund, R. E. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting", **European Conference On Computational Learning Theory**, Springer-Verlag London, UK, 23-37, 1995.
- [37] Internet: OpenCV Documentation, Cascade Classification: Haar Feature-based Cascade Classifier for Object Detection, http://docs.opencv.org/2.4/modules/objdetect/doc/cascade_classification.html, 14.03.2017.
- [38] D. Hefenbrock, et al., "Accelerating Viola-Jones Face Detection to FPGA-Level using GPUs", **18th IEEE Annual International Symposium on Field-Programmable Custom Computing Machines**, 2010.
- [39] H. Jia, Y. Zhang, et al., "Accelerating Viola-Jones Face Detection Algorithm On GPUs", **IEEE 14th International Conference on High Performance Computing and Communications**, 2012.
- [40] A.W.Y. Wai, S.M. Tahir, and Y.C. Chang, "GPU Acceleration of Real Time Viola-Jones Face Detection", **IEEE International Conference on Control System, Computing and Engineering**, 2015.
- [41] L. Sun, S. Zhang, et al. "Acceleration Algorithm for CUDA-based Face Detection", **IEEE International Conference on Signal Processing, Communication and Computing (ICSPCC 2013)**, 2013.
- [42] C. Sagonas et al., "300 Faces in-the-Wild Challenge: The first facial landmark localization Challenge", **International Conference on Computer Vision**, 2013.
- [43] P. N. Belhumeur, D. W. Jacobs, D. J. Kriegman, and N. Kumar, "Localizing parts of faces using a consensus of exemplars.", **2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)**, 545–552, 2011.
- [44] V. Le, J. Brandt, Z. Lin, L. Bourdev, and T. S. Huang, "Interactive facial feature localization.", **Computer Vision–ECCV 2012**, 679–692, 2012.
- [45] X. Zhu and D. Ramanan, "Face detection, pose estimation, and landmark localization in the wild.", **2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)**, 2879–2886, 2012.
- [46] M. Kostinger, P. Wohlhart, P. M. Roth, and H. Bischof, "Annotated facial landmarks in the wild: A large-scale, realworld database for facial landmark localization.", **2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops)**, 2144–2151, 2011.
- [47] Internet: 300 Faces In-The-Wild Challenge (300-W), IMAVIS 2014, https://ibug.doc.ic.ac.uk/resources/300-W_IMAVIS/, 14.03.2017.
- [48] D. E. King, "Dlib-ml: A Machine Learning Toolkit", *J Mach Learn Res.*, 10, 1755-1758, 2009.
- [49] Internet: Dlib: Image Processing Shape Predictor, http://dlib.net/imaging.html#shape_predictor, 14.03.2017.
- [50] N. Cristianini, J.S. Taylor, "An Introduction to Support Vector Machines and other kernel based learning methods", *Cambridge University Press*, 2001.
- [51] U. Ayvaz, M. Peker, H. Gürüler, "Uykuda periyodik bacak hareketleri bozukluğunun bir karar destek sistemi ile tespiti", **Uluslararası Mühendislik Teknolojileri ve Uygulamalı Bilimler Konferansı ICETAS 2016**, Afyon Kocatepe Üniversitesi, Afyon, 1138-1141, 2016.
- [52] K. P. Soman, R. Loganathan, V. Ajay, "Machine learning with SVM and other kernel methods", PHI Learning Pvt. Ltd., New Delhi, India, 2009.
- [53] Ü. Aydoğan, "Destek vektör makinelerinde kullanılan çekirdek fonksiyonların sınıflama performanslarının karşılaştırılması", Yüksek Lisans Tezi, Hacettepe Üniversitesi, Sağlık Bilimleri Enstitüsü, 2010.
- [54] Internet: M. Bernstein, The Radial Basis Function Kernel, <http://pages.cs.wisc.edu/~matthewb/>, 14.03.2017.
- [55] Internet: Orange Data Mining Tool, <http://orange.biolab.si>, 14.03.2017.
- [56] S. Datta, D. Sen, R. Balasubramanian, "Integrating Geometric and Textural Features for Facial Emotion Classification Using SVM Frameworks.", **Proceedings of International Conference on Computer Vision and Image Processing. Advances in Intelligent Systems and Computing**, 459, 619-628, 2016.
- [57] A. Basu, A. Routray, S. Shit, and A. K. Deb, "Human emotion recognition from facial thermal 253 image based on fused statistical feature and multi-class svm.", **2015 Annual IEEE India Conference 254 (INDICON)**, 1–5, 2015.
- [58] R. Niese et al., "Facial expression recognition based on geometric and optical flow features in colour image sequences", *IET Computer Vision*, 6(2), 79-89, 2012.
- [59] H. Candra et al., "Classification of facial-emotion expression in the application of psychotherapy using Viola-Jones and Edge-Histogram of Oriented Gradient.", **2016 IEEE 38th Annual International Conference of the Engineering in Medicine and Biology Society (EMBC)**, 423-426, 2016.