



**Trakya Üniversitesi
Mühendislik Bilimleri Dergisi**

Cilt: 17 Sayı: 2 Aralık 2016

**TRAKYA
UNIVERSITY
JOURNAL OF
ENGINEERING
SCIENCES**

Volume: 17 Number: 2 December 2016

Trakya Univ J Eng Sci

<http://dergipark.gov.tr/tujes>
tujes@trakya.edu.tr

ISSN 2147-0308

**Trakya Üniversitesi
Mühendislik Bilimleri Dergisi**

Cilt: 17

Sayı: 2

Aralık

2016

**Trakya University
Journal of Engineering Sciences**

Volume: 17

Number: 2

Aralık

2016

Trakya Univ J Eng Sci

<http://dergipark.gov.tr/tujes>
tujes@trakya.edu.tr

ISSN 2147-0308

Dergi Sahibi / Owner

Trakya Üniversitesi Rektörlüğü, Fen Bilimleri Enstitüsü Adına
On behalf of Trakya University Rectorship, Graduate School of Natural and Applied Sciences
Prof. Dr. Murat YURTCAN

Baş Editör / Editor-in-Chief

Doç. Dr. Hacı Ali GÜLEÇ

Dizgi / Design

Hayriye KAHVECİOĞLU, h.kahvecioglu7@gmail.com

İletişim Bilgisi / Contact Information

Address : Trakya Üniversitesi, Enstitüler Binası, Fen Bilimleri Enstitüsü,
Balkan Yerleşkesi, 22030, Edirne / TÜRKİYE

Web site : <http://dergipark.gov.tr/tujes>

E-mail : tujes@trakya.edu.tr

Tel : +90 284 2358230

Fax : +90 284 2358237

Baskı / Publisher

Trakya Üniversitesi Matbaa Tesisleri
Trakya University Publishing Centre

Editör Kurulu / Editorial Board

Altan MESUT	Bilgisayar Mühendisliği Bölümü	Trakya Üniversitesi
Ayşegül AKDOĞAN EKER	Makine Mühendisliği Bölümü	Yıldız Teknik Üniversitesi
Aysu UĞURLAR	Şehir ve Bölge Planlama Bölümü	Yüzüncü Yıl Üniversitesi
Aytaç ALPASLAN	Elektrik-Elektronik Mühendisliği Böl.	Trakya Üniversitesi
A. Can ZÜLFİKAR	İnşaat Mühendisliği Bölümü	Trakya Üniversitesi
Burhan ÇUHADAROĞLU	Makine Mühendisliği Bölümü	Karadeniz Teknik Üniversitesi
Cem S. ÇETİNARSLAN	Makine Mühendisliği Bölümü	Trakya Üniversitesi
Esmâ MIHLAYANLAR	Mimarlık Bölümü	Trakya Üniversitesi
Gökhan KAÇAR	Genetik ve Biyo-mühendislik Bölümü	Trakya Üniversitesi
İsa CAVİDOĞLU	Gıda Mühendisliği Bölümü	Yüzüncü Yıl Üniversitesi
Metin AYDOĞDU	Makine Mühendisliği Bölümü	Trakya Üniversitesi
Mustafa ERGEN	Kentsel Tasarım ve Peyzaj Mim. Böl.	Amasya Üniversitesi
Özer GÖKTEPE	Tekstil Mühendisliği Bölümü	Namık Kemal Üniversitesi
Pelin ONSEKİZOĞLU BAĞCI	Gıda Mühendisliği Bölümü	Trakya Üniversitesi
Rukiye Duygu ÇAY	Peyzaj Mimarlığı Bölümü	Trakya Üniversitesi
Semra HASANÇEBİ	Genetik ve Biyo-mühendislik Bölümü	Trakya Üniversitesi
Timur KAPROL	Mimarlık Bölümü	Trakya Üniversitesi
Tolga SAKALLI	Bilgisayar Mühendisliği Bölümü	Trakya Üniversitesi
Tülay YILDIRIM	Elektronik ve Haberleşme Müh. Böl.	Yıldız Teknik Üniversitesi
Türkan GÖKSAL ÖZBALTA	İnşaat Mühendisliği Bölümü	Ege Üniversitesi
Utku GÜNER	Biyoloji Bölümü	Trakya Üniversitesi
Ümit GEÇGEL	Gıda Mühendisliği Bölümü	Namık Kemal Üniversitesi

İÇİNDEKİLER / CONTENTS

An application for the failure mode and effects analysis integrated with the grey relational analysis

Burak KURŞUN, Ünal KURT, Salih GÜVERCİN, Korhan ÖKTEN, Simge AKGÜL, Aytaç YILDIZ 35-44

Genetic algorithm and differential evolution algorithm compared on a novel application domain

Didem ABİDİN 45-56

Çift zamanlı ve yarı yapılı özel veri ambarları ile ilişkisel veri tabanı üzerine bir karşılaştırma çalışması

Gözde ALP, Canan Eren ATAY 57-68

AN APPLICATION FOR THE FAILURE MODE AND EFFECTS ANALYSIS INTEGRATED WITH THE GREY RELATIONAL ANALYSIS

Burak KURŞUN¹, Ünal KURT², Salih GÜVERCİN³, Korhan ÖKTEN¹, Simge AKGÜL⁴ Aytaç YILDIZ^{5*}

¹ Department of Mechanical Engineering, Amasya University, Amasya-TURKEY

² Electric-Electronic Engineering, Amasya University, Amasya-TURKEY

³ Machine Technology Program, Amasya University, Amasya-TURKEY

⁴ Cumhuriyet University, Koyulhisar MYO, Sivas-TURKEY

⁵ Department of Automotive Engineering, Amasya University, Amasya-TURKEY

Abstract: Although the Failure Mode and Effects Analysis (FMEA) is a systematic method of analysis, it has some shortcomings and limitations since it is a method based on intuitionistic and subjective statements of the person that rate the failure modes. In order to eliminate these constraints, the use of the method in conjunction with the grey relational analysis, which is one of the multi criteria decision making methods, helps to eliminate intuitionistic situations and prioritize the failure modes that need corrections and precautions.

The classical FMEA and the FMEA integrated with the grey relational analysis approaches were applied, and their effectiveness was assessed in this study to identify and prioritize the failures and determine the measures to be taken in the wheat sieving machine production. For this purpose, first the Risk Priority Numbers (RPN) were calculated using the classical failure mode and effects analysis, then two separate grey RPNs were calculated on the assumptions that risk factors have either equal weight or different weight in the grey relational analysis-integrated FMEA approach, and the prioritization of the failures was performed. Three different RPN values obtained in the study were compared, and the priority optimizations to be made were recommended in order to prevent failures before reaching the customers as well as drawing the necessary conclusions accordingly.

Keywords: Failure Mode and Effects Analysis; grey relational analysis; multi criteria decision making

GRİ İLİŞKİSEL ANALİZİ İLE BÜTÜNLEŞTİRİLMİŞ HATA TÜRÜ VE ETKİLERİ ANALİZİ YAKLAŞIMI İÇİN BİR UYGULAMA

Özet: Hata Türü ve Etkileri Analizi (HTEA) sistematik bir analiz yöntemi de olsa hata türlerini değerlendiren kişilerin subjektif ifadelerine dayanması sebebiyle aynı zamanda sezgisel de bir yöntemdir. Sezgilere dayanması ise uygulamada bazı eksikliklere ve kısıtlamalara yol açmaktadır. Bu problemleri ortadan kaldırabilmek için yöntemi çok kriterli karar verme yöntemlerinden biri olan gri ilişkisel analizi ile birlikte kullanmak, sezgisel durumları ortadan kaldırarak önlem alınmasını ve düzeltilmesi gereken hata türlerinin önceliklendirilmesini sağlamaktadır.

Bu çalışmada buğday eleme makinesi üretimindeki hataların tespit edilerek önceliklendirilmesi ve alınacak önlemlerin belirlenmesi için klasik HTEA ve gri ilişkisel analizi ile bütünleştirilmiş HTEA yaklaşımları uygulanarak yaklaşımların etkinliği değerlendirilmiştir. Bunun için ilk önce klasik HTEA ile Risk Öncelik Sayıları (RÖS) daha sonra ise gri ilişkisel analizi ile bütünleştirilmiş HTEA yaklaşımıyla risk faktörlerinin hem eşit ağırlığa hem de farklı ağırlıklara sahip olduğu varsayımıyla iki ayrı gri RÖS hesaplanarak hataların önceliklendirilmesi yapılmıştır. Çalışma sonunda elde edilen üç ayrı RÖS değerleri karşılaştırılmış ve buna göre hataların müşteriye ulaşmaması için öncelikle yapılması gereken iyileştirmeler önerilmiş ve gerekli değerlendirmeler yapılmıştır.

Anahtar Kelimeler: Hata Türü ve Etkileri Analizi; gri ilişkisel analizi; çok kriterli karar verme

Introduction

Before the market launch of a new product, many problems arise in the design, pre-production planning, production, packaging, shipping, and the customer presentation stages. Identifying and resolving this problem in the relevant processes before the product reaches to the customer is of great importance for the company. Some failures can be determined easily, whereas some others cannot be detected, and remain hidden eventually. These hidden remaining failures can only be unearthed by monitoring and evaluating the process intensively as well as using various quality management techniques. The FMEA method is one of the most effective methods used for identifying the factors that may cause a problem in the process regarding the product. Although the FMEA method is a systematic method, there may be intuitive situations in the determination and evaluation of the failures. The integration of the method with mathematical models will reduce the emergence of intuitive situations (Down et al., 2008). For this purpose, both the classical FMEA and the FMEA integrated with the grey relational analysis approaches were applied in three separate models in order to prioritize the failure modes, with the assumptions that risk factors have either equal or different weights, and the effectiveness of three models were evaluated within the scope of detection and prioritization of the failures that may arise in the wheat sieving machine production, which is in the product range of a company that produces agricultural machinery in Turkey.

In the following second section of the study, the FMEA method is briefly described, in the third section, the grey relational analysis integrated with FMEA approach is described, the application on the company's case is performed in the fourth section, and the conclusions are drawn in the final section in accordance with the results obtained.

FMEA Method

The FMEA method was developed by the American army, and the first procedure prepared for this purpose is MIL-P-1629 (Military Procedure), dated November 9, 1949, titled as 'Procedures for Performing a Failure Mode, Effects, and Criticality Analysis'. The method was first used for the evaluation of system and equipment failures (Gulcicek and Sofyalioglu, 2014), then it was adapted in areas that produce safety-critical products that contain advanced electronic and mechanical equipment, in advanced manufacturing systems, such as automotive and aerospace systems (Baysal et al., 2002). In 1975, it has been implemented the manufacture of computers and in Japanese NEC company. It became widespread in the automobile industry, particularly Ford in 1977 and Fiat in 1985, and also applied by Chrysler, and General Motors as well (Aytac, 2011).

The definition of FMEA in MIL-STD-1629A (Procedures for Applying the Failure Mode and Effects Analysis) published in 1980, which is one of the first published standards on this topic, is given as "A procedure to analyze and classify each possible failure mode in the system according to their significance, and determining their outcomes and effects in the system." Stamatis (1995) expands the definition, and states that "FMEA is an engineering method that aims to detect, identify, and eliminate the known and/or potential failures, mistakes and problems in the design, process, systems and services before they reach the customer" (Buyuktuna, 2012; Kadioglu, et al. 2009; Cevik and Aran, 2009). According to another definition, FMEA is an analytical method that is used to identify potential problems that may arise in the product or process development works and is the most explicit documentation of the information that is collected by cross-functional teams (Down et al., 2008). FMEA focuses on the potential failures that may occur in order to locate and eliminate the root causes of the

product or process before they reach the customer. It improves the operational performance of the production cycle, reduces the risk by performing system analysis or collecting historical data (Scipioni et al., 2002; Omdahl, 1988). Potential risks and impacts of each failure in the plant equipment are listed in the order of criticality on the system. In other words, this method examines the equipment, malfunctioning situations, and potential impacts (Yucel, 2007; Dizdar and Kurt, 1996). With this method, the likelihood of failures is reduced, to produce quality products or services that can respond to customers' needs and expectations (Buyuktuna, 2012). In addition, FMEA helps the competitiveness of the company, improves company image, reduces product development time and cost, helps in the development of new production methods, and allows the reduction of scrap and waste (Eryurek and Tanyas, 2003). The FMEA method is divided into the system FMEA, design FMEA, process FMEA, and the service FMEA classifications (Chin et al., 2008).

In FMEA, the Risk Priority Number (RPN) is calculated first by multiplying three important factors (Eq.1). These factors are the occurrence (*O*), severity (*S*), and detection (*D*) respectively.

$$RPN=O \times S \times D \tag{1}$$

The occurrence refers to the frequency of the occurrence of the risk. The severity factor refers to the impact if that risk happens. The detection factor is a measure of the detectability of the risk before it occurs. The occurrence, severity, and detection values used in FMEA are presented in Tables 1, 2, and 3 respectively (Wang et al., 2009).

Table 1. Crisp Ratings for Occurrence of a Failure (Wang et al., 2009).

Rating	Probability of occurrence	Failure probability
10	Very high: failure is almost inevitable	>1 in 2
9		1 in 3
8	High: repeated failures	1 in 8
7		1 in 20
6	Moderate: occasional failures	1 in 80
5		1 in 400
4		1 in 2000
3	Low: relatively few failures	1 in 15.000
2		1 in 150.000
1	Remote: failure is unlikely	<1 in 1.500.000

Table 2. Crisp Ratings for Severity of a Failure (Wang et al., 2009).

Rating	Effect	Severity of effect
10	Hazardous without warning	Very high severity ranking when a potential failure mode effects safe system operation without warning
9	Hazardous with warning	Very high severity ranking when a potential failure mode effects safe system operation with warning
8	Very high	System inoperable with destructive failure without compromising safety
7	High	System inoperable with equipment damage
6	Moderate	System inoperable with minor damage
5	Low	System inoperable without damage
4	Very low	System operable with significant degradation of performance
3	Minor	System operable with some degradation of performance
2	Very minor	System operable with minimal interference
1	None	No effect

Table 3. Crisp Ratings for Detection of a Failure (Wang et al., 2009).

Rating	Detection	Likelihood of detection
10	Absolute uncertainty	Design control cannot detect potential failure mode
9	Very remote	Very remote chance, the design control will detect potential failure mode
8	Remote	Remote chance, the design control will detect potential failure mode
7	Very low	Very low chance, the design control will detect potential failure mode
6	Low	Low chance, the design control will detect potential failure mode
5	Moderate	Moderate chance, the design control will detect potential failure mode
4	Moderately high	Moderately high chance, the design control will detect potential failure mode
3	High	High chance, the design control will detect potential failure mode
2	Very high	Very high chance, the design control will detect potential failure mode
1	Almost certain	Design control will detect potential failure mode

The flow of the FMEA process is shown in Figure 1.

The FMEA approach integrated with the grey relational analysis

The grey theory, which has been developed by Julong Deng in 1982 for the first time, is widely used to solve problems of uncertainty under missing or incomplete information. In addition, the grey theory is a popular method that is used in decision making in multi-criteria cases as well as analyzing various relationships between discrete data sets (Sofyalioglu, 2011).

The results based on the original data, simple and easily understandable calculations, and being the best

method for decision making are the main advantages of this method (Wu, 2002).

Chang et al. (2001) have suggested that the grey theory can be applied to FMEA, since the risk factors in classical FMEA have the characteristics mentioned above. Grey relational analysis is one of the subheadings of the grey theory. This analysis

technique is used to determine the degree of relationship between the series of reference factors and each factor in a grey system. Each factor is defined as a series (rows or columns), and the degree of effect between factors is called as the grey relational degree (Sofyalioglu, 2011).

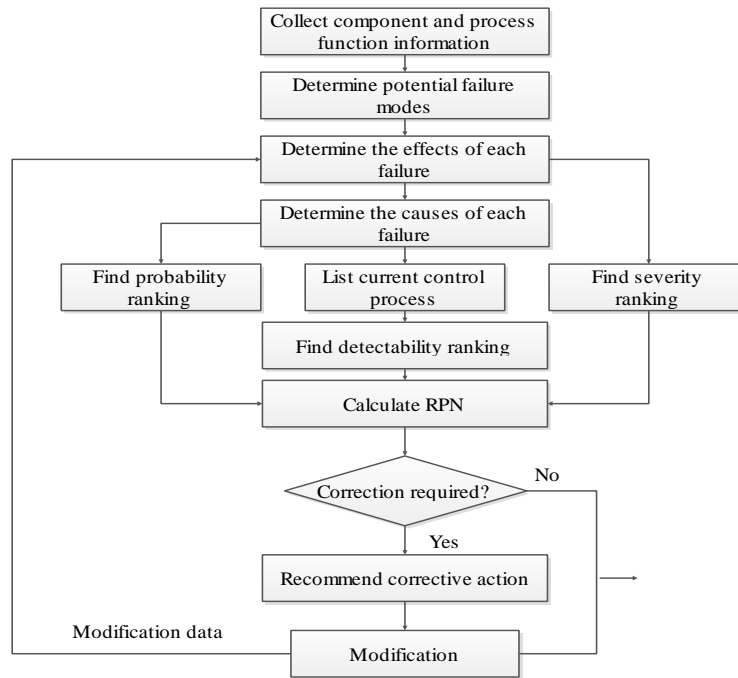


Figure 1. The FMEA process (Pillay and Wang, 2003).

The FMEA Approach Integrated with the grey Relational Analysis is performed with the following steps (Chang et al., 2001):

Step 1. Determination of standard (reference) series: target series, known as the standard series, of length k is as follows

$$X_0 = \{X_0(1), X_0(2), \dots, X_0(k)\} \quad (2)$$

Since small scores indicate smaller risks, standard series are determined by taking the smallest values of all risk factors.

$$X_0 = \{X_0(1), X_0(2), \dots, X_0(k)\} = \{1, 1, \dots, 1\} \quad (3)$$

Step 2. Determination of comparative series: m information series related to the occurrence, severity, and detection of the failure can be expressed as in the

following.

$$X_i = \{X_i(1), X_i(2), \dots, X_i(k)\} \quad i = 1, 2, \dots, m \quad (4)$$

$X_i(k)$; shows the k th factor of X_i . If it is possible to compare all information series, n information series can be defined as in the following matrix.

$$X = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{bmatrix} = \begin{bmatrix} X_1(1) & X_1(2) & \dots & X_1(k) \\ X_2(1) & X_2(2) & \dots & X_2(k) \\ \vdots & \vdots & \ddots & \vdots \\ X_n(1) & X_n(2) & \dots & X_n(k) \end{bmatrix} \quad (5)$$

Step 3. Obtaining the difference between comparative series and standard series, and calculating the coefficient value: the difference between the decision factor point. Standard series model is determined to reveal the degree of fuzzy relation. All three risk factors are compared with the standard series for each failure series, and the relation coefficient is defined by

the following equation;

$$\gamma(X_0(k), X_j(k)) = \frac{\Delta_{min} + (\zeta \Delta_{max})}{\Delta_{0j}(k) + (\zeta \Delta_{max})} \quad (6)$$

Here $j=1, \dots, m$ and $k=1, \dots, n$ $X_0(k)$; standard series. $X_j(k)$: comparative series

$$\Delta_{min} = \min_{j \in I} \min_{k \in K} \|X_0(k) - X_j(k)\| \quad (7)$$

$$\Delta_{max} = \max_{j \in I} \max_{k \in K} \|X_0(k) - X_j(k)\| \quad (8)$$

ζ is defined as a coefficient between (0, 1) . In order to reduce the effect of the maximum value on the relationship coefficient, this coefficient is taken as $\zeta = 0.5$ usually. The purpose of this coefficient is to adjust the difference between " $\Delta_{0j}(k)$ " and " Δ_{max} ". Studies show that ζ value does not affect the order obtained as a result of the grey relational degree.

Step 4. Determination of the grey relational degree: the grey relationship degree is a measure of the geometric similarity between comparative series and standard (reference) series. The degree of the relationship indicates that there is a strong relationship between the comparative and standard series.

If the compared two series are the same, then the grey relationship degree is found as 1. The grey relationship degree shows the extent of similarity between compared and standard series.

$$\tau(X_0, X_j) = \frac{1}{n} \sum_{k=1}^n \gamma(X_0(k), X_j(k)) \quad (9)$$

n is the decision factor number.

If the weights of each criteria are given, the grey relationship degree can be found by multiplying the weight values of the degree significance of the criterion and grey relationship coefficient of the criterion.

$$\tau(X_0, X_j) = \sum_{k=1}^n \gamma(X_0(k), X_j(k)) W_k \quad (10)$$

Here W_k is the weighting coefficient of the factors, and

$$\sum_{k=1}^n W_k = 1 \quad (11)$$

In the classical FMEA, the occurrence, severity, and detectability of the failures cannot be assigned appropriately to reflect the real world. The grey approach in FMEA, however, allows engineers to assign relative weights to the decision factors based on research and production strategies.

Step 5. Rating the risk priorities: in decision making problems, when the standard (reference) series is selected as the ideal series, then the grey relationship degree, which is calculated between the compared series and reference series, will be an indicator of the degree of detectability of the criteria. In other words, the factor series with the highest grey relational degree indicate the best alternative in decision making problems (Sofyalioglu, 2011). That is if $\gamma(X_0, X_i) \geq \gamma(X_0, X_j)$ we can say that the relationship between X_i and X_0 is greater than the relationship between X_j and X_0 . For FMEA, this shows the relationship between the scores of potential causes and optimal values of decision factors. Greater degree of relationship indicates the lesser effect of the source of failure. Therefore, gradually increasing the degree of relationship shows the decrease in risk priorities of the potential causes to be improved (Sofyalioglu, 2011; Chang et al., 2001).

The case study

The study was conducted with the data obtained from a company, operating in the manufacturing sector in Turkey that produces agricultural machinery. The wheat sieve machine, which is one of the products of the company, is a machine having 4 sieves used to sieve wheat. The company wants to reduce costs and increase customer satisfaction by determining the failures, arisen during the design and production of the machine before they reach the customers. For this reason, the grey relational analysis-integrated FMEA and the classical FMEA methods that detect, prioritize

and allow to taking precautions were implemented to this company.

Classical FMEA Application

In the first part of the study, the classical FMEA approach was applied to detect, prioritize, and eliminate failures that occur in the design and production of the wheat sieving machine. In the first phase of implementation, a team of 4 people was established with individuals from the design, production, quality control and marketing departments

of the company in order to determine the potential failure modes, causes, and effects of these failures. This team has determined the potential failures, effects, and causes of the failures given in Table 4 through brainstorming and using cause-and effect diagrams, and the RPN values were calculated with the help of Eq.1.

When we examine Table 4, it is seen that the failure numbers 1, 2, and 3 have RPN values 640, 600, and 480 respectively, indicating the priority failures that need to addressed first.

Table 4. Results of evaluation of classical FMEA.

Identifying the Design and Production Failure Modes of a Four-Sieve Wheat Sieving Machine Using the FMEA Method								
CURRENT SITUATION								
Failure No	Potential Failure	Effect of Potential Failure	Severity	Potential Causes of the Failure	Occurrence	Preventive Control in the Present Process	Detectability	RPN
1	Failure of suction fan	Inability to perform a complete cleaning	8	Insufficient hood	8	High amounts of air pressure	10	640
2	Frequent breakdown and breakage of the parts due to vibrations	Inoperative machinery	10	Welding defects and thin pieces used	6	Replacement or re-welding of the broken parts	10	600
3	Breakage of the brush arms	Inoperative brushes	6	Random bending of the brush arms	8	Replacement of the brush arms	10	480
4	Wood hangers used in the assembly of sieves	Inoperative machinery	3	Wooden hanger	10	Replacement of the broken part	10	300
5	Loosening of the fixing screws due to poor floor of customer's site and high vibration of the machine	Excessive vibration, malfunctioning and stopping of the machine as a result of the loosening of the screws	10	Poor floor at the installation site	4	When loosen, they have screwed again until it's not possible to screw further, then liquid concrete is poured and fixed again.	10	400
6	The holes do not align since the protection covers are pre-drilled before the installation.	Increased cost and waste of time during installation	3	Pre-drilled holes of the cover before installation	8	New ones are made with larger holes	10	240

Application of the FMEA Approach Integrated with the Grey Relational Analysis

In this approach, the RPN values calculated above

using the classic FMEA method are converted to grey RPN values by using the grey relational analysis and re-prioritized. According to this new order of priority,

the priority failures are determined, and the problems that cause these failures are eliminated. The calculation of grey RPN values of the failure modes, calculated by applying the steps of the grey-relational analysis integrated FMEA approach, and the prioritization of failures are as follows.

First, the following information series matrix is obtained by using the RPN values given in Table 4 in order to determine the comparative series.

$$\begin{bmatrix} X_1(1) & X_1(2) & X_1(3) \\ X_2(1) & X_2(2) & X_2(3) \\ X_3(1) & X_3(2) & X_3(3) \\ X_4(1) & X_4(2) & X_4(3) \\ X_5(1) & X_5(2) & X_5(3) \\ X_6(1) & X_6(2) & X_6(3) \end{bmatrix} = \begin{bmatrix} 8 & 8 & 10 \\ 10 & 6 & 10 \\ 6 & 8 & 10 \\ 3 & 10 & 10 \\ 10 & 4 & 10 \\ 3 & 8 & 10 \end{bmatrix}$$

Then, the information series matrix is used to calculate the following difference matrix.

$$\begin{bmatrix} \Delta_1(1) & \Delta_1(2) & \Delta_1(3) \\ \Delta_2(1) & \Delta_2(2) & \Delta_2(3) \\ \Delta_3(1) & \Delta_3(2) & \Delta_3(3) \\ \Delta_4(1) & \Delta_4(2) & \Delta_4(3) \\ \Delta_5(1) & \Delta_5(2) & \Delta_5(3) \\ \Delta_6(1) & \Delta_6(2) & \Delta_6(3) \end{bmatrix} = \begin{bmatrix} 7 & 7 & 9 \\ 9 & 5 & 9 \\ 5 & 7 & 9 \\ 2 & 9 & 9 \\ 9 & 3 & 9 \\ 2 & 7 & 9 \end{bmatrix}$$

According to the values of the difference matrix, $\Delta_{min}=2$, $\Delta_{max}=9$ was defined, and the determination coefficient (ζ) was accepted as 0.5. After obtaining the difference matrix, the grey relationship coefficients were calculated according to Equation 6. For $\Delta_1(1)$;

$$\gamma(X_0(k), X_i(k)) = \frac{\Delta_{min} + (\zeta \Delta_{max})}{\Delta_{01}(1) + (\zeta \Delta_{max})} = \frac{2 + (0.5 \times 9)}{7 + (0.5 \times 9)} = 0.565 \quad (12)$$

was calculated. The same procedure was performed for the others to find the grey relationship coefficients in the following matrix.

$$\begin{bmatrix} \gamma_1(1) & \gamma_1(2) & \gamma_1(3) \\ \gamma_2(1) & \gamma_2(2) & \gamma_2(3) \\ \gamma_3(1) & \gamma_3(2) & \gamma_3(3) \\ \gamma_4(1) & \gamma_4(2) & \gamma_4(3) \\ \gamma_5(1) & \gamma_5(2) & \gamma_5(3) \\ \gamma_6(1) & \gamma_6(2) & \gamma_6(3) \end{bmatrix} = \begin{bmatrix} 0.565 & 0.565 & 0.481 \\ 0.481 & 0.684 & 0.481 \\ 0.684 & 0.565 & 0.481 \\ 1.000 & 0.481 & 0.481 \\ 0.481 & 0.867 & 0.481 \\ 1.000 & 0.565 & 0.481 \end{bmatrix}$$

In the final stage, grey relationship degree (grey RPN) of each failure mode was calculated for determining

the risk priorities. At this stage, first of all, the grey relationship degrees were calculated using Equation 9, under the assumption that all three risk factors have equal weights. For example, the grey relationship degree in relation to 1st failure mode was calculated as;

$$\tau(X_1, X_3) = \frac{1}{3} (0.565 + 0.565 + 0.481) = 0.537 \quad (13)$$

The same process was applied to other failure modes to obtain;

$$\text{Grey RPN} = \begin{bmatrix} 0,537 \\ 0,549 \\ 0,577 \\ 0,654 \\ 0,609 \\ 0,682 \end{bmatrix}$$

Then, the grey relationship degrees of the failure modes were re-calculated using Equation 10 under the assumption that the risk factors have different weights. By taking $\sum_{k=1}^n W_k = 1$, the weight coefficients for the occurrence, detectability, and severity of the failure were taken $W_O=0,4$, $W_D=0,2$ ve $W_S=0,4$ respectively. For example, the grey relationship degree in relation to 1st failure mode was calculated as;

$$\tau(X_1, X_3) = (0.565 \times 0.4 + 0.565 \times 0.2 + 0.481 \times 0.4) = 0.531 \quad (14)$$

The same process was applied to other failure modes to obtain;

$$\text{Grey RPN} = \begin{bmatrix} 0,531 \\ 0,520 \\ 0,578 \\ 0,688 \\ 0,557 \\ 0,705 \end{bmatrix}$$

At the end of the application, RPNs, grey RPNs for the risk factors having the same and different weight for the failure modes, and the priority ranks of the failure modes are shown in Table 5.

When we examine Table 5, it is seen that priority ranks of the failure modes calculated with the classical FMEA, and grey relational analysis-integrated FMEA

with and without equal weights of risk factors were the same. In either case, the most significant two failures that need to be addressed were "Failure of suction fan", and "Frequent breakdown and breakage of the parts due to vibrations" respectively. The two most significant failure modes in the other two cases (classical FMEA and grey RPN that risk factors have equal weight) have exchanged ranks here, and the

"Frequent breakdown and breakage of the parts due to vibrations" now became the most significant failure mode that needs to be addressed.

After these comparisons and findings, the improvements in relation to failure modes in Table 6 were proposed to the company, and the company has addressed these recommendations to keep failures from reaching the customer.

Table 5. RPN values.

Failure Modes	RPN	Ranking	Grey RPN (Risk factors have equal weight)		Grey RPN (Risk factors have different weight)	
				Ranking		Ranking
1	640	1	0.537	1	0.531	2
2	600	2	0.549	2	0.520	1
3	480	3	0.577	3	0.578	4
4	300	5	0.654	5	0.688	5
5	400	4	0.609	4	0.557	3
6	240	6	0.682	6	0.705	6

Table 6. Improvements on failure modes.

Proposed Improvements	Authorized Person	Actual Improvements	Severity	Occurrence	Detectability	RPN
Changing hood design to have wider holes	Assembler	Large air holes were designed.	2	2	3	12
Making the thin machine parts 2 mm thicker	Welder, designer	2 mm thicker parts were produced to replace the broken parts.	3	2	3	18
Performing the brush arm bending in certain molds.	Designer, producer	A custom mold was made for brush arms.	3	4	2	24
Use of metal hanger in the assembly of sieves	Designer, producer	The metal hanger was used.	3	3	3	27
Improving the quality of the concrete for mounting the machine at the customer	Customer, designer	The quality of the ground was improved.	4	2	3	24
Not drilling the holes of machine protection covers before installation	Designer, assembler	Protective covers were drilled after the installation of the machine	2	2	2	8

Conclusions

The FMEA method aims at preventing failures from reaching the customer by taking corrective measures and detecting the failures in the design, and production of a product or service. However, the classical FMEA may not provide reliable results

in the prioritization of the failures since it is open to subjective assessments in the determination of risk factors as well as in the detection, occurrence, and severity of the failures. Therefore, the grey relational analysis-integrated FMEA approach was implemented in order to eliminate the

disadvantages of the classical FMEA method, and compared with the results obtained from the classical FMEA method.

The priority ranks of the failure modes were the same in the classic FMEA method and the grey relational analysis-integrated FMEA approach applied under the assumption that the risk factors have equal weights. However, when we assumed that the risk factors do not have equal weights in accordance with the opinions of the team established in the company, four of the failure modes changed ranks. What is important here is the change in the ranking of the first two failure modes that need to be addressed. Considering the failure modes under the assumption that risk factors have different weights, the "Frequent breakdown and breakage of the parts due to vibrations" failure mode was found to be more critical since it affects operation of the machine, whereas the "Failure of suction fan" failure mode was only affecting the complete cleaning, hence it was determined that failure in the operation of the machine is more significant failure and needs immediate precaution. As understood from this result, the classical FMEA and the assumption that risk factors have equal weight may give erroneous results in the priority rankings due to the restrictions previously described. However, in situations where risk factors have different weights gave more meaningful results, and this result is in line with the study by Sofyalioglu (2011).

In future studies, grey RPN values can be calculated on the assumption that risk factors have different weights and using the improvement data obtained by the classical FMEA to compare the results.

References

AYTAÇ E. Kalite iyileştirme sürecinde bulanık mantık yaklaşımı ile hata türü ve etkileri analizi ve

uygulama örneği. *Adnan Menderes Üniversitesi Sosyal Bilimler Enstitüsü Doktora Tezi*. 2011.

BAYSAL ME., CANIYILMAZ E., EREN T. Otomotiv yan sanayinde hata türü ve etkileri analizi. *Teknoloji Dergisi*. 5(1-2): 83-90, 2002.

BÜYÜKTUNA O. Hata türü ve etkileri analizi ve makine sanayinde bir uygulama. *Pamukkale Üniversitesi Sosyal Bilimler Enstitüsü Yüksek Lisans Tezi*. 2012.

CHANG CL., LIU PH., WEI CC. Failure mode and effect analysis using grey theory. *Integrated Manufacturing Systems*. 12(3): 211-216, 2001.

CHIN KS., CHAN A., YANG JB. Development of a fuzzy FMEA based product design system. *International Journal of Advanced Manufacturing Technology*. 36: 633-649, 2008.

ÇEVİK O., ARAN G. Kalite iyileştirme sürecinde hata türü etkileri analizi (FMEA) ve piston üretiminde bir uygulama. *SÜ İİBF Sosyal ve Ekonomik Araştırmalar Dergisi*. 16: 242-265, 2009.

DİZDAR EN., KURT M. Olası iş kazaları için bir erken uyarı modeli. *Gazi Üniversitesi Fen Bilimleri Enstitüsü Dergisi*. 9(4): 525-532, 1996.

DOWN M., BROZOWSKI L., YOUNIS H., BENEDICT D., FEGHALI J., SCHUBERT M., BRENDER R., GRUSKA G., VALLANCE G., KRASICH M., HAUGHEY W. Potential failure mode and effect analysis (FMEA) reference manual (4th Ed.), Chrysler LLC, Ford Motor Company, General Motors Corporation, 2008.

ERYÜREK ÖF., TANYAŞ M. Hata türü ve etkileri analizi yönteminde maliyet odaklı yeni bir karar verme yaklaşımı. *İTÜDERGİSİ/d*. 2(6): 31-40, 2003.

GÜLÇİÇEK B., SOFYALIOĞLU Ç. Bulanık kalite fonksiyon göçerimi ile hata türü ve etkileri analizinin bir ambalaj firmasında uygulanması. *Yönetim*. 21(2): 73-97, 2014.

KADIOĞLU M., UÇMUŞ E., GÖNEN D. Makine imalatı yapan bir işletmede tasarım hata türü ve etkileri analizi ile hata kaynaklarının belirlenmesi ve kalitenin iyileştirilmesi. *BAÜ FBE Dergisi*. 11(1): 42-55, 2009.

OMDAHL TP. Reliability, availability and maintainability dictionary, American Society for Quality, Quality Press, Milwaukee, United States of America, 1988.

PILLAY A., WANG J. Modified failure mode and effects analysis using approximate reasoning. *Reliability Engineering and System Safety*. 79: 69-85, 2003.

SCIPIONI A., SACCAROLA G., CENTAZZO A., ARENA F. FMEA methodology design, implementation and integration with HACCP system in a food company. *Food Control*. 13: 495-

501, 2002.

SOFYALIOĞLU Ç. Süreç hata modu etki analizini gri değerlendirme modeli. *Ege Akademik Bakış*. 11(1): 155-164, 2011.

STAMATIS D.H. Failure mode and effects analysis: FMEA from theory to execution, ASQ Quality Press, 1995.

WANG YM., CHIN KS., POON GKK., YANG JB. Risk evaluation in failure mode and effects analysis using fuzzy weighted geometric mean. *Expert systems with applications*. 36(2): 1195-1207, 2009.

WU, H.H. A Comparative study of using grey relational analysis in multiple attribute decision making. *Quality Engineering*. 15(2): 209-217, 2002.

YÜCEL Ö. Konfeksiyon üretiminde hata türü ve etkileri analizi. *Tekstil ve Konfeksiyon*. 2: 126-131, 2007.

GENETIC ALGORITHM AND DIFFERENTIAL EVOLUTION ALGORITHM COMPARED ON A NOVEL APPLICATION DOMAIN

Didem ABİDİN*

Celal Bayar University, Department of Computer Engineering, 45140 Manisa, Turkey

Abstract: In this paper, the performances of genetic algorithms (GA) and differential evolution (DE), which are two of the most popular optimization techniques used, are compared. There exist many other studies which compare these two; however, comparing those on an education material domain will be the contribution to the literature. The problem is stated as a sequencing problem of education material, in which the order of the topics covered really matters. Selection of the contents of the courses to be given to the students is an important factor to improve the level of education of the students. Representing the content of a course in the correct order is a critical task for instructors. In this study, the importance of the order of the contents of a course was emphasized and the performance of a course content sequencing mechanism using GA and DE was compared. The results put forward that, sequencing the course contents with GA performs better; however, DE is also obviously successful with a close score to that of the GA's.

Keywords: Genetic algorithm; differential evolution; sequencing; optimization

GENETİK ALGORİTMA VE DİFERANSİYEL EVRİM ALGO-RİTMASININ YENİ BİR UYGULAMA ALANINDA KARŞILAŞTIRILMASI

Özet: Bu çalışmada, genetik algoritmalar (GA) ve diferansiyel evrim (DE) algoritmaları gibi çok popüler iki optimizasyon tekniği kullanılarak performans karşılaştırmaları yapılmıştır. Bu iki tekniği kıyaslayan pek çok çalışma yapılmış olsa da, bu kıyaslamaların algoritmaların eğitim materyalleri üzerinde kullanılarak yapılmış olması literatüre katkı değerinde olacaktır. Ele alınan problem bir sıralama problemi olup, işlenen konuların sırası önem kazanmaktadır. Ders içeriğinin doğru seçimi, öğrencilerin eğitim seviyesinin yükselebilmesi için çok önemlidir. Dersin öğretim üyesi için, ders içeriğini doğru sıra ile aktarmak kritik bir görevdir. Bu çalışmada, bir ders içeriğinin doğru sırada aktarılmasının önemine dikkat çekilmiş ve GA ile DE tekniklerinin kullanılmasıyla oluşturulmuş ders içeriği sıralama mekanizmasının performans karşılaştırmaları yapılmıştır. Alınan sonuçlar, GA'nın DE tekniğinden biraz daha iyi sonuçlar elde ettiğini göstermiştir.

Anahtar Kelimeler: Genetik algoritmalar; diferansiyel evrim; sıralama; optimizasyon

Introduction

Artificial intelligence (AI), which is a branch of computer science, aims to understand “intelligence” by developing some computer programs which simulates the behavior of an intelligent being. Accordingly, computer science education at universities has become very popular. Many different courses organized by computer science departments of universities have emerged and these courses have to be arranged according to the needs of the current technological developments.

The selection of computer science courses is important but determining the contents of a course is also crucial.

The contents of a course have to be up to date; they have to meet the requirements of the sector. The contents of a course have to be helpful for the students when they graduate and start working as computer scientists. For this purpose, the instructors have to decide the order of the contents of a course because finding the optimum order for the contents makes learning easier for the students. Although this is a decision that the instructor has to make, a computer program can help to find an optimum order for the course content more quickly and more reliably. Genetic algorithm (GA) is a stochastic search technique which can be used in various application domains from optimization (Rankovic, 2014)

to sequencing problems. Many studies in literature emphasize that the GA parameters have an effect on the performance of sequencing problems. Route planning problems can be considered as this kind of problem and GA was used successfully in Wu et al. (2009) with a standard one-point crossover operator. Order acceptance problems can be classified as another branch of sequencing problems, in which a two-point crossover technique can be applied (Rom and Slotnick, 2009). Aircraft departure sequencing problems is also a popular search topic, where re-searchers have developed many techniques to solve it (Wang et al., 2009). They have used a reliable range of parameter values like the crossover rate range of 0.70-1.00, which is also the chosen range for this paper. Studies which search for solutions to different optimization problems have also chosen mutation and crossover rates (Hsieh et al., 2009), as chosen in this paper. Curriculum sequencing is also handled in de Marcos et al. (2008) for different courses having different features from each other, which can be considered as a rule based approach.

Differential evolution (DE) algorithm is also a powerful optimization tool which was developed by Price and Storn in 1995 and its structure is similar to that of GA (Storn and Price, 1995). Although DE is a tool for continuous optimization, it also can be used for some combinatorial optimization problems like Travelling Salesman Problem (TSP) (Prado et al., 2010). It works population based as well as GA. In the population, there exist some possible solutions for the problem to be solved. DE uses a heuristic optimization technique to find the best solution among the candidates (Keskinturk, 2006; Storn and Price, 1995). The common aspect of GA and DE is using similar genetic material transfer methods like crossover and mutation as in the nature. Even so, usage of the genetic operators in GA and DE is different. GA works crossover based and mutation is applied in 1-1.5% of the population. Unlike GA, DE applies the mutation operation to each individual while transferring them to the next generation. In its mutation

operation, for each individual, three more individuals are selected from the population and as the result of the mutation operation, a mutant individual is obtained. Fitness value of the mutant individual is calculated. According to the fitness value, it is decided whether to be replaced with the first individual chosen or not. Crossover is not the main operation in DE as it is in GA.

In this study, the performances of two different methodologies for the solution of sequencing education material problem are compared. With the software implemented, sequencing is done automatically by a computer program with a genetic algorithm method in it, which was used in previous studies (Abidin and Cakır, 2011; Abidin and Cakır, 2014). The reason for choosing the course material is to emphasize that educational material can also be optimized by using certain evolutionary algorithms, which makes the study a computer science application. Database course (DB) and algorithms course (ALG), which are compulsory for computer science students, are chosen for the experimental tests of the project. A comparison is made between the reliability performances of GA and DE. Different crossover and mutation rates will be used in the software as Hegerty et al. (2009) did, which also compared GA and DE. For both GA and DE, best results are also compared with the suggestions of an expert, who is the instructor of the course, as aforementioned in (Abidin and Cakır, 2014) via the Spearman Rank Correlation Test. The Spearman correlation test is used to prove the reliability of decision support systems (Balli et al., 2009) and in solving optimization problems with GA (Garcia-Camacho et al., 2011) and DE.

What make the sequencing difficult for this kind of problem are the interrelations among the contents of the courses. Some of the content may be shuffled but some of it cannot be replaced. These interrelations are stated as a set of prerequisite rules within the content. These rules are stated and saved both as logical rules and mathematical representations to make the computation easier. This makes the study an effective solution to a

precedence constrained sequencing problem (Yun, Gen, Moon, 2010). GA and DE use these sets of rules while deciding the order of the course content.

The layout of the paper is as follows: Section 2 includes the problem definition, characteristics of data and the solution analysis, while Section 3 discusses the system output and covers the interpretations of results from different points of view. The discussion concerning the evaluation results and future work appear in Section 4.

Material and Method

Among all computer science courses, the contents to be sequenced are chosen as the contents of a Database course (DB) and an Algorithms (ALG) course. There are many topics in those courses which have to be sequenced. These selected contents are determined by the instructor of the course via a user interface. Each part of the contents of the course is called a “module” . This format can easily be applied to some other educational material for different courses because educational content of every course can be divided into small sections and these sections should be instructed with a certain order. The modules to be sequenced have to obey some prerequisite rules; they must be in the correct order according to their antecedents. When a module *i* has the module *j* as its prerequisite, this means that module *j* has to be represented before module *i*. The module names of both the Algorithms and Database courses are given in Table 1 and Table 2 respectively.

For Database course, 17 modules were chosen by the instructor to be sequenced. For Algorithms course, 20 modules were chosen. GA and DE algorithms were applied to DB and ALG modules respectively (with different values of crossover and mutation rates as given in Hegerty et al. (2009)) and the best module ranges for each method were obtained. Both GA and DE are run for fixed generation values and the best individuals at the end of the last generation are taken into consideration for the evaluation phase of the study.

Table 1. Modules for algorithms course.

Module Name	Prerequisites
1 – Iterations (while)	6-3-14-8-19
2 – Executing a Program	6-10-14-8-20-19
3 – Constants	6-10
4 – Console I/O Operations	6-10-3-14-12-8-20-19-2
5 – Matrices	6-3-14-12-8-19-1-7-15
6 – Introduction to Algorithms	-
7 – Numerical Problems	6-3-14-12-8-20-19-7
8 – Operators	6-3-14-12
9 – Bit Operations	6-14-12-8-19
10 – Programming Environments	6
11 – String Operations	6-14-12-8-20-19-15
12 – Type Conversion and Casting	6-14
13 – Subroutines	6-10-3-14-12-8-20-19-18-2-4
14 – Variables	6-10-16-3
15 – Arrays	6-3-14-12-8-19-1-7
16 – Comments	10
17 – Switch/Case	6-10-14-12-8
18 – Break/Continue	6-3-14-8-19-1
19 – Iterations (for)	6-3-14-12-8-20
20 – If/Then/Else	6-10-3-14-12-8

Table 2. Modules for database course.

Module Name	Prerequisites
1 –Introduction to DBMS	-
2 – DBMS-DBA	1
3 – Data Models	1
4 – Table	1-2
5 – Referential Integrity	1-2 -4
6 – Algebraic Operators	1-2-4
7 – PK	1-3-4 -5
8 – FK	1-3-4-5-7
9 – Normalization	1-4-5-7-8
10 – DB Design	1-4-5-7-9
11 – Data Types	3-4-10
12 – SQL-Select	3-4-5-7-8-11
13 – SQL-Insert	3-4-5-7-8-11-12
14 – SQL Operators	4-12-13
15 – SQL-Update	3-4-5-7-8-11-12-14
16 – SQL-Delete	3-4-5-7-8-11-12-14
17 – Sequences	4-7-8

Both of the results from GA and DE were compared with the expert’ s suggestions using a nonparametric correlation test as done in the previous study (Abidin and Çakır, 2014). Like Roeva (2008), each different

parameter combination was executed for 20 times and the average generation number was obtained (Hegerty et al., 2009). The same tests were repeated for 6 different population sizes of DB and 2 different population sizes of ALG. Because of the structural differences of GA and DE, no mutation values are used in DE. Instead, DE has a different parameter F, which is used as the constant value in mutation equations. Table 3 shows the values of parameter combinations of GA and DE.

Table 3. GA and DE parameters.

Parameter	Value (GA)	Value (DE)
Population Size	100, 120, 140, 160, 180, 200	50, 100
# of Generations (G)	500, 750, 1000	1000, 2000, 3000
Crossover Rate (crate)	0.7, 0.75, 0.8, 0.85, 0.9, 0.95, 1	0.5, 0.8
Mutation Rate (mrate)	0.1, 0.15, 0.2	-
F value	-	0.5, 0.6

Both sets of GA and DE results are also compared with the expert's suggestion by performing tests with the Spearman Rank Correlation to decide their reliability.

GA Features

As Holland (1975) stated, GA is a simulation of the natural reproduction process of living things to transfer their genetic knowledge to the next generations.

This idea has been embraced by other scientists like Goldberg (1989), and has been improved by many others and applied to the solutions of many recent daily life problems. Since the chromosomes of GA represent one possible solution to the module sequencing problem, each gene of the chromosome consists of different module numbers. For this reason, permutation encoding is the most appropriate representation method for this project. The initial population is obtained from the random sequences of the modules given in previous section.

Three different crossover operators were applied in four different ways. That is; 1-point order crossover (called

GA1), 2-point crossover (called GA2), partially mapped crossover-PMX.

As the mutation operator, swap mutation is applied. Linear Rank Selection (Grefenstette and Baker, 1989) was used as the selection mechanism of the population. The two individuals with the best fitness values are transferred to the next generation directly in order to apply elitism.

The fitness function of a GA is the only part which has to be designed according to the needs and structure of the problem. In this study, the fitness function mechanism works by parsing the 1s in the prerequisite matrix and the penalty scores are calculated for the modules on the chromosome. This is quite a practical way of transferring the rules to a mathematical representation. Saving the rules in a matrix has a few benefits. The system does not have to execute any queries from the database to acquire the rule data. The equation for the fitness function is given in Eq. 1.

$$fitness[i] = \frac{1}{\sum_{k=1}^m \frac{penaltycount_k}{prereqcount_k} + 1} \quad (1)$$

where m is the number of modules in a chromosome, *penaltycount* is the number of prerequisite modules for module k existing in the chromosome and *prereqcount* is the total number of prerequisite modules for module k. After the penalty scores are calculated, the individual with the lowest penalty point is determined as the individual with the best fitness value. The individual with the best fitness value, more likely, is supposed to transfer its genetic material to the next generation.

DE Features

The population on which DE was run consists of individuals (chromosomes) with randomly sequenced module numbers. For ALG, the number of genes in each chromosome of the population is 17. All individuals are updated and transferred to the next generation until the determined number of generations (G) is reached. Like GA, DE also needs to use fitness functions for the

individuals throughout generations in order to understand whether better individuals are generated or not. The same fitness mechanism, depending on minimizing the prerequisites penalties, was used in DE as well as in GA.

After calculating the fitness values of individuals, the individuals and their fit-ness values are sorted from the worst individual to the best. In this way, the best individual at the end of G generations is always the last individual of the popula-tion.

In a standard DE algorithm, a mutant individual is generated by choosing some of the individuals from the population. Since the data set of this study is not suitable for repeating genes or real number representation, a DE version which was used in TSP is preferred (Greffenstette and Baker, 1989). Accordingly, four different mutation strategies are applied: MX1: DE/rand/1 (Ghosh et al., 2012), MX2: DE/best/1 (Storn and Price, 1997), MX3: Simplex1 (Kamiyama et al., 2010), MX4: Simplex2 (Kamiyama et al., 2010).

In all of these strategies, a mutant individual is created by the help of randomly chosen individuals and the best and the worst individuals of the population. The formulas to create the mutant individual are given for MX1, MX2, MX3 and MX4 in Equation 1, Equation 2, Equation 3 and Equation 4 respectively.

In DE/rand/1 strategy:

Except the chosen individual (x_i, g), three more individuals are chosen randomly (x_1, x_2, x_3). For each gene of the three individuals, the mutant individual is calculated (Equation 2). Mutant individual is shown with ($v_{i, g+1}$).

For each gene in the chromosome ($n=0..19$ for ALG):

$$v_{i, g+1}[n] = x_1[n] + F * (x_2[n] - x_3[n]) \quad (2)$$

In DE/best/1 strategy:

Except the chosen individual (x_i, g), two more individuals are chosen randomly (x_1, x_2). For the

best individual of the population and x_1 and x_2 , the mutant individual is calculated (Eq. 3).

For each gene in the chromosome ($n=0..19$ for ALG) and NP given as the number of individuals in the population:

$$v_{i, g+1}[n] = \text{Pop}[\text{NP}-1][n] + F * (x_1[n] - x_2[n]) \quad (3)$$

In Simplex1 strategy:

Except the chosen individual (x_i, g), one more individual is chosen randomly (x_1).

For the best and the worst individuals of the population and for x_1 , the mutant individual is calculated (Eq. 4).

For each gene in the chromosome ($n=0..19$ for ALG) and NP given as the number of individuals in the population:

$$v_{i, g+1}[n] = x_1[n] + F * (\text{Pop}[\text{NP}-1][n] - \text{Pop}[0][n]) \quad (4)$$

In Simplex2 strategy:

Except the chosen individual (x_i, g), one more individual is chosen randomly (x_1).

For the best, second best and the worst individuals of the population and for x_1 , the mutant individual is calculated (Eq. 5).

For each gene in the chromosome ($n=0..19$ for ALG) and NP given as the number of individuals in the population:

$$v_{i, g+1}[n] = x_1[n] + F * (\text{Pop}[\text{NP}-1][n] - \text{Pop}[0][n] + \text{Pop}[\text{NP}-2][n] - \text{Pop}[0][n]) \quad (5)$$

Repair Mechanism

Since DE is preferred in solving continuous optimization problems (Agrawal and Srikant, 1994), repairing mechanism is needed to be added to DE for this study. In all applied strategies, the genes of the mutant individuals have values in real numbers. Real numbers cannot be used to represent the values of the

genes in this study; therefore a conversion operation is done on the genes of the chromosomes (Mi et al., 2010) and they are converted to module numbers of ALG. For the same conversion operation, Relative Position Indexing approach can also be used which is suggested by Prado et al. (2010). The resulting mutant individual is used in the crossover phase.

In crossover, the mutant individual ($v_{i,g+1}$) and the individual previously selected from the population ($x_{i,g}$) are used. A random value between 0 and 1 is generated and it is compared with the already determined CR value. If the random value is greater than CR, mutant individual is marked as the candidate individual ($u_{i,g}$) to the next generation. If the random value is less than CR, $x_{i,g}$ is marked to be transferred to the next generation as ($u_{i,g}$). Crossover operation is given in Eq. 6 (Elsayed et al., 2012).

$$u_{i,g} = \begin{cases} v_{i,g+1} \rightarrow rand[0,1] > CR \\ x_{i,g} \rightarrow rand[0,1] \leq CR \end{cases} \quad (6)$$

A fitness value is calculated for the candidate individual ($u_{i,g}$). This value is compared with that of the first chosen individual ($x_{i,g}$). If the fitness value of the candidate individual is greater, that individual is transferred to the next generation of the population instead of ($x_{i,g}$). If the fitness value of the candidate individual is not greater, then the original individual ($x_{i,g}$) again takes place in the next generation of the population.

Reliability of Sequences

Since many result sets were obtained for both DB and ALG, these result sets had to be examined to find the most reliable result when compared with the expert's suggestion. When a judgment is to be made on a group of data which is in sequences, it is convenient to use the Spearman Rank Correlation. It is a nonparametric test used in statistical analysis. It is used in cases where testing the reliability of the range of data is more important than the numerical values of the data (Sheskin, 2000). With this test, a t value is calculating

for each sequence and a reliability percentage can be obtained. Equation (7) shows the calculation of t values.

$$t = \frac{\rho}{\sqrt{(1-\rho^2)/(n-2)}} \quad (7)$$

where ρ indicates the Spearman rank correlation coefficient and n is the number of alternatives (observations). The number of observations was 17 for DB and 20 for ALG. Then the hypothesis should be set as given below:

H0: There is no correlation between the two ranges.

H1: There is a correlation between the two ranges.

The evaluation criteria for the Spearman test are given in Table 4 below.

Table 4. Spearman correlation test parameters.

Number of Modules (n)	17	20
Degree of Freedom (n-2)	15	18
Tolerance	1%	
Confidence Level (p) (two tailed)	0.005	0.005
T value	2.947	2.878

The reliability of the module ranges were evaluated with a tolerance percentage of 1% ($P < 0.01$). According to the hypothesis, to be able to understand whether there is a correlation between the output and the expert's suggestion, the t values must be compared with a value of 2.947 for DB and with a value of 2.878 for ALG according to the t table used for this study (Bissonette, 2010). To verify the reliability of the modules, the p values should also be calculated. If the calculated p value is smaller than the tolerance percentage ($P < 0.01$) then the hypothesis H0 is rejected and it can be stated that the module sequence is reliable.

Findings

The results of GA were examined according to the following criteria: the runtime values in minutes, t values, reliability percentages and the number of reliable results. According to the results, the execution time for GA2 is by far the best among other GAs for all population sizes. This is because the performance of 2-point order crossover is much better than the

performances of 1-point crossover and PMX. In Table 5, the reliability percentages found with GA are given for DB.

Table 5. Reliability percentages of results for DB (GA)

Pop. Size	Reliability Percentages		
	GA1	GA2	PMX
100	95.83	96.81	95.10
120	96.56	98.04	94.36
140	95.83	97.30	96.08
160	97.06	97.30	95.59
180	96.08	98.53	94.36
200	95.34	96.81	96.07

In the results of DB, a population size of 180 with 2-point order crossover gives the most reliable results with a percentage of 98.53. For the best results, the two-tailed p value for these ranges is less than 0.0001 ($P < 0.01$); therefore, both ranges are accepted as extremely significant.

Since Spearman Rank Correlation was used in computing fitness values in DE, the best individuals of the populations of result sets are already the ones having the highest reliability percentages. Table 6 shows the reliability percentages of the module ranges of ALG with DE according to their population sizes when compared with the expert's suggestion (Zhang et al., 2008).

Table 6. Reliability percentages of results for ALG (DE)

Pop. Size	Reliability Percentages			
	De/rand/1	De/best/1	Simplex1	Simplex2
50	91.12	92.63	94.28	92.78
100	88.72	91.57	96.84	94.43

In the results of ALG, a population size of 100 with Simplex1 mutation gives the most reliable results with a percentage of 96.84. For the best results, the two-tailed p value for these ranges is less than 0.0001 ($P < 0.01$); therefore, both ranges are accepted as extremely significant.

If the best results for 50 individuals and 1000 generations are examined, it can be seen that these results are obtained for the same crossover rate (CR=0.5)

and F value (F=0.5). For 2000 generations, the best results are gathered for the same crossover rate and F values as gathered for 1000 generations. If the algorithm is run for 3000 generations, the best results are obtained with the Simplex2 strategy, again with the same crossover rate of 0.5.

If the best results for 100 individuals and 1000 generations are examined, the best results are obtained with Simplex1 strategy. Although crossover rates and F values of the best results does not follow a significant pattern, three strategies except "De/Best/1" have the same crossover rate of 0.5.

For 2000 generations, Simplex2 strategy has the best results. Simplex1 strategy has closer results to the best and both strategies have the same crossover and F values (F=0.5, CR=0.5). The best F value for all four strategies is observed as 0.5.

If the number of generations is increased to 3000, best results are observed for Simplex1 strategy. Except for "De/rand/1" strategy, most reliable results are gathered for the crossover rate CR=0.5 and F=0.5.

The number of successful solutions was calculated according to the t value formula of the Spearman Rank Correlation. The calculated values were then compared with the corresponding value in t-table (Bissonette, 2010). The corresponding t value of t-table to be compared is detected according to the Spearman Correlation Test parameters used in this study. The software implemented also gave the best t value of all 1260 runs of GA. Having the highest t value indicates that the result giving the highest t value is the best solution among the result set. Mostly the best t values of different population sizes are from the operator GA2. This shows that 2-point order crossover gives predominantly the best module range results. The numbers of reliable results are given in Table 7.

Table 7. The number of reliable results for DB (out of 1260)

Pop. Size	GA1	GA2	PMX
100	967	1148	785
120	991	1188	840
140	1049	1218	914
160	1053	1231	948
180	1085	1238	986
200	1114	1253	1050

99.4% of the results are found as reliable for the 200 individual dataset and 2-point crossover. When the corresponding crate values of the successful scenarios are examined in GA, it can be observed that the number of scenarios with 0.95 as the crossover rate is higher than the other crossover rates.

The same extraction can be done for the mutation rates. The number of scenarios with the mutation rate of 0.1 was greater than the others in the dataset. Higher mutation rates did not perform well for DB.

Being the best solution was not the only answer which this study looks for. Be-sides the best result, the number of all reliable results should be taken into consideration. For each GA and DE and for each population size, different numbers of reliable solutions have been acquired. The numbers of reliable results for DE in ALG are given in Table 8.

Table 8. The number of reliable results for ALG (out of 240)

Pop. Size	De/rand/1	De/best/1	Simplex1	Simplex2
50	159	139	214	233
100	197	192	236	240

With Simplex2, 100% of the results are found as reliable. From the obtained re-liable sequence amounts' point of view, Simplex1 and Simplex2 strategies have the highest number of reliable sequences for the population value of 100. When the population is set as 100 individuals, Simplex2 has completely reliable results (240 out of 240 for all numbers of generations). In DE, 96 different result sets are obtained for 4 different mutation strategies, 2 different F values, 2 different

crossover rates, 2 different population values and 3 different generations. Each combination was run for 20 times; which means that software was run for 240 times for each mutation strategy and population size in total. This number is less than that of the GA's, but the reason is that, in DE there are no mutation rates as in GA and there are less crossover rates. In DE tests, the crossover rate of 0.5 has always given the best results. Therefore, this rate can be used as a fixed value for DE to be used for course material optimization. For the F value, F=0.5 has given the best result in most of the tests. The results obtained from Simplex1 and Sim-plex2 strategies are obviously better than the ones obtained by "De/rand/1" and "De/best/1" .

Results

In the guidance of the previous studies, it was investigated whether DE could be used for optimizing the sequence of course material. While applying DE to the dataset, four different mutation strategies are tested, as done before with GA (Abidin and Çakır, 2011) (Abidin and Çakır, 2014). To be able to make meaningful comparisons, the results of optimization of the course material with GA was also included to the study. The population size interval for GA was given as 100 – 200 (100, 120, 140, 160, 180 and 200), where it was 50 and 100 for DE. The system needs to test different sizes of populations because increasing the population size to a certain extent encourages the diversity of the population that both GA and DE deal with. This means that as the population size increases, the possibility of having individuals with various values of fitness also increases. Choosing different numbers of populations for GA and DE was done intention-ally because GA and DE react to different numbers of populations. Increasing the number of populations in GA gave better results; however, increasing the num-ber of generations gave better results in DE.

It is very important to verify that the study stands as a reliable one among all other scientific research in the

relevant field literature. In this study, the output of the software is compared with the solution suggested by an instructor to decide the curriculum planning system's reliability. For this reason, all of the results obtained in the study with all scenarios and genetic operators are accepted as the best solutions at the beginning of the evaluation process and the Spearman Rank Correlation test was applied to all.

From the genetic operators' point of view, in GA, keeping the crossover rates higher gave better results (0.95) both in 1 - point order and 2 - point order crossover operators. The mutation rates tested were extremely high in this study when compared with similar studies in literature, though the preliminary tests of the study indicated that lower mutation rates like 0.01, 0.02, etc. did not give better results in this kind of sequencing problem.

The results suggest that, the rule-based GA developed in this study can be used as a reliable system to optimize a curriculum sequence with tight prerequisite rules among sections of educational material. This system is quite suitable for use in sequencing the contents of all courses given in a computer science or computer engineering department of a faculty. To verify the practicality of this study, a whole content sequencing tool and its implementation for a specific department in a faculty like computer engineering is planned as a future study.

The best result with DE was obtained with a reliability of 96.84% (Simplex1 strategy, NP=100). The best result with GA was obtained with a reliability of 98.53% (2-point crossover, NP=180). This shows that this software application (with both GA and DE) can be reliably helpful to introduce the course materials to the students.

GA and DE will be used in a different application domain as the future work, with the experience gathered from this study.

References

ABIDIN, D., ÇAKIR, Ş. (2011): Rule-Based Genetic Algorithm for In-Service Training Curriculum Plan. Proc International Conference on Computers, Digital Communications and Computing (ICDCCC ' 11), Barcelona, Spain:160 - 166.

ABIDIN, D., ÇAKIR, Ş. (2014): Analysis of a Rule Based Curriculum Plan Optimization System with Spearman Rank Correlation. Turkish Journal of Electrical Engineering and Computer Sciences, 22:176-190.

AGRAWAL, R., SRIKANT, R. (1994): Fast Algorithms for Mining Association Rules. Proc 20th International Conference on Very Large Databases, Morgan Kaufmann, Santiago, Chile:487-499.

BALLI, S., et al. (2009): Neuro - Fuzzy Decision Support System for Selecting Players in Basketball. Journal of İstanbul Technical University, 8:15-25.

BISSONNETTE, V.L. (2011): Critical Values of the t Distribution. [Online]: <http://facultyweb.berry.edu/vbissonnette/tables/t.pdf>

DE MARCOS, L., et al. (2008): Competency-Based Curriculum Sequencing: Comparing Two Evolutionary Approaches. Proc IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology, Sydney, Australia:339 - 342.

ELSAIED, S.M., SARKER, R.A., ESSAM, D.L. (2012): On an Evolutionary Approach for Constrained Optimization Problem Solving. Applied Soft Computing, Elsevier, 12:3208-3227.

GARCIA-CAMACHO, F., et al. (2011): Genetic Algorithm Based Medium Optimization for a Toxic Dinoflagellate Microalga. Harmful Algae, 10:697-701.

GHOSH, S., et al. (2012): A Differential Covariance Matrix Adaptation Evolutionary Algorithm for Real Parameter Optimization. Information Sciences 182:199-219, Elsevier.

- GOLDBERG, D.E. (1989): Genetic Algorithms in Search, Optimization and Machine Learning. Reading, Addison Wesley.
- GREFFENSTETTE, J.J., BAKER, J.E. (1989): How Genetic Algorithms Work: A Critical Look at Implicit Parallelism. Proc 3rd International Conference on Genetic Algorithms, San Mateo, CA, USA, Morgan Kaufmann Publishers:20-27.
- HEGERTY, B., HUNG, C.C., KASPRAK, K. (2009): A Comparative Study on Differential Evolution and Genetic Algorithms for Some Combinatorial Problems. 1st Workshop on Intelligent Methods in Search and Optimization - WIMSO.
- HOLLAND, H.J. (1975): Adaptation in Natural and Artificial Systems. Cambridge: MIT Press.
- HSIEH, S.-T., SUN, T.-Y., LIU, C.-C., (2009): Potential Offspring Production Strategies: An Improved Genetic Algorithm for Global Numerical Optimization. International Journal of Expert Systems with Applications, 36:11088-11098.
- KAMIYAMA, D., TAMURA, K., YASUDA, K. (2010): Down-hill Simplex Method Based Differential Evolution. Proc SICE Annual Conference, Taipei, Taiwan:1641-1646.
- KESKINTÜRK, T. (2006): Differential Evolution Algorithm. İstanbul Trade University Journal of Science, 9: 85-99.
- MI, M., et al. (2010): An Improved Differential Evolution Algorithm for TSP Problem. International Conference on Intelligent Computation Technology and Automation:544-547.
- PRADO, R.S., et al. (2010): Using Differential Evolution for Combinatorial Optimization: A General Approach. Proc IEEE International Conference on Systems, Man and Cybernetics, Istanbul, Turkey:11-18.
- RANKOVIC, V., et al. (2014): The Mean-Value at Risk Static Portfolio Optimization Using Genetic Algorithm. Computer Science and Information Systems 11(1):89-109.
- ROEVA, O. (2008): Improvement of Genetic Algorithm Performance for Identification of Cultivation Process Models. Proc 9th WSEAS International Conference on Evolutionary Computing (EC' 08), Sofia, Bulgaria:34-39.
- ROM, W.O., SLOTNICK, S.A. (2009): Order Acceptance Using Genetic Algorithms. Computers & Operations Research Elsevier Ltd., 36:1758-1767.
- SHESKIN, D.J. (2000): Handbook of Parametric and Non-Parametric Statistical Procedures. 2nd Ed. Chapman & Hall / CRC.
- STORN, R., PRICE, K. (1995): Differential Evolution: A Simple and Efficient Adaptive Scheme for Global Optimization over Continuous Spaces. Technical Report TR-95-012, International Computer Science Institute, Berkeley.
- STORN, R., PRICE, K. (1997): Differential Evolution: A Simple and Efficient Adaptive Scheme for Global Optimization over Continuous Spaces. Journal of Global Optimization 11:341-359.
- WANG, L.-J., HU, D.-W., GONG, R.-Z. (2009): Improved Genetic Algorithm for Aircraft Departure Sequencing Problem. Proc 3th IEEE International Conference on Genetic and Evolutionary Computing:35-38.
- WU, M.-C., SHIH, C.-F., CHEN, C.-F. (2009): An Efficient Approach to Cross - Tab Route Planning for Water Manufacturing. International Journal of Expert Systems with Applications 36:11962-11968.
- YUN, Y., GEN, M., MOON, C. (2010): Hybrid Genetic Algorithm with Adaptive Local Search for Precedence-Constrained Sequencing Problems. Proc 40th IEEE Computers and Industrial Engineering Conference (CIE' 10):1-6.
- ZHANG, J., et al. (2008): Differential Optimization for

Discrete Optimization: An Experimental Study on Combinatorial Auction Problems. IEEE Congress on Evolutionary Computation (CEC'08), Hong Kong.

ÇİFT ZAMANLI VE YARI YAPILI ÖZEL VERİ AMBARLARI İLE İLİŞKİSEL VERİ TABANI ÜZERİNE BİR KARŞILAŞTIRMA ÇALIŞMASI

Gözde ALP^{*1}, Canan Eren ATAY²

¹ Marmara Üniversitesi, Mühendislik Fakültesi, Bilgisayar Mühendisliği Bölümü, İstanbul, Türkiye

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

Özet: Veri ambarları ve ilişkisel veri tabanları büyük veriyi saklamak ve işlemek için kullanılan iki farklı depolama alanıdır. Bu çalışmada; yapısal olarak özelleştirilmiş iki zamanlı ve iç içe yapıda bir veri ambarı hazırlanmıştır. Aynı veriler veri ambarında ve ilişkisel veri tabanında saklanarak bu iki depolama alanı arasında karşılaştırma yapılmıştır. Aynı sonucu verecek sorgular hem veri tabanı tabloları kullanılarak hem de veri ambarı bileşenleri kullanılarak sorgulanmıştır. Sonuçlar sorgu setleri, yürütme zamanları ve disk alanı kullanımı açısından değerlendirilmiştir ve karşılaştırılmıştır.

Anahtar Kelimeler: Veri ambarları; büyük veri; öznitelik zaman etiketi; çift zamanlı veri tabanları

A COMPARATIVE STUDY ON BITEMPORAL SEMISTRUCTURED DATA WAREHOUSES AND RELATIONAL DATABASES

Özet: Data warehouses and relational databases are two different areas for data storage. A special structured data warehouse is prepared in this study that is bitemporal and its structure is formed as nested. Same data is stored in data warehouse and the relational database. These two data storage area is competed with each other. Same resulting queries are queried by using relational database objects and data warehouse components. The results are compared and evaluated in terms of query sets, processing time and disk space usage.

Keywords: Data warehouse; big data; bitemporal databases; attribute time stamping.

Giriş

Günümüzde teknolojinin etkileri her sektörde görülmektedir. Sağlıktan, tekstile, eğitimden, yiyecek içecek sektörüne kadar her alanda, bilgiler bilgisayar ortamında kayıt altında tutulmakta ve işlenmektedir. Bunun yanı sıra, insan vücudunun bir uzantısı haline gelen akıllı cep telefonları ile insanlar sosyal ortamlarda kişisel bilgilerini, beğenilerini paylaşır hale gelmiştir. Bilgi paylaşımında güvenilir yöntemlerin geliştirilmesiyle ve veri paylaşım hızının artmasıyla, özellikle son yıllarda internet üzerinden eşya ve yiyecek alışverişi dünyada ve ülkemizde insan hayatının sıradan bir parçası haline gelmiştir. Bilgisayarsız bir hayat düşünülemez hale gelmiştir.

Bilgisayar kullanımı bu denli yaygın değilken kayıt altında tutulamayan birçok veri zamanımızda kayıt altında tutulabilmektedir. Ortaya çıkan büyük veri akışı

verileri işleme ve bu verileri firmaların ve toplumların yararına kullanma ihtiyacı ortaya çıkmasına sebep olmuştur. Öyle ki artık şirketler tüketici eğilimlerini anlamak ve karlarını arttırmak adına ellerindeki verileri işleyecek ve akılcı önerilerde bulunabilecek firmalar ile çalışmaktadır. Dahası bazı büyük şirketlerin veri işleme bölümleri bulunmaktadır. Bunun yanında hastalık teşhis ve tedavilerini geliştirebilmek için sağlık alanında da büyük veri topluma faydalı olması için kullanılmaktadır.

Bu ihtiyaç doğrultusunda veri işleme ve veriden bilgi elde edebilme üzerinde çalışan bilimler ortaya çıkmıştır. Karar destek sistemleri, veri madenciliği yöntemleri büyük veriyi anlamlandırmak için ortaya çıkmış bilimsel alanlardır. Veri ambarları ise yüksek boyuttaki veriyi saklamak ve işlemek için elverişli bir depolama alanlarıdır. Şirketlerin büyük bir çoğunluğu veri ambarında sakladıkları verileri karar destek sistemleri

ya da veri madenciliği yöntemleri ile işleyerek ortaya çıkan bilgiyi kullanarak karar almaktadır.

Akla gelen her sektörde ve sosyal platformda ortaya çıkan bu veriler ilişkisel veri tabanlarında ya da veri ambarlarında saklanabilir. Günümüzde; Microsoft, Oracle gibi şirketler tarafından ya da açık kaynak kodlu olarak üretilmiş onlarca değişik ilişkisel veri tabanı ve veri ambarı sistemleri mevcuttur.

İlişkisel veri tabanı; her an veri ekleme, silme, update etme gibi işlemlerin yapılabilirdiği çevrimiçi olarak adlandırabileceğimiz veri depolama alanıdır. Veri ambarında ise çevrimiçi işlem yapılmaz, sadece sorgulama, raporlama işlemleri yapılır. Veri belli zamanlarda ilişkisel veri tabanı ya da benzer veri depolarından veri ambarına aktarılır.

Sahip olunan eski, önceki yıllara dönemlere ait veriler, gelecek dönemlerdeki stratejileri belirlemekte kullanılabilir. Veri ambarları özel yapıları ile geçmiş dönem analizinin yapılmasını kolaylaştırır. Veri ambarı içerisinde bulunan veriler zamansal derinliğe sahiptir. Veri ambarları, büyük miktarda verinin saklanabileceği bir depolama alanı sunarken aynı zamanda çevrimiçi işleminin olmadığı sadece sorgulama işlemleri yapılan ve dolayısıyla daha hızlı sonuç veren de bir ortam sağlar.

Veri ambarı; eski verilere ulaşmak ve analiz etmek amacı ile kullanılan, aktif bir çalışma alanıdır. Veri ambarı “yöneticilerin karar vermelerini kolaylaştıran; konu odaklı, entegre, kalıcı, zamanla değişen verilerin oluşturduğu bir yığındır” (Inmon ve Krishnan, 2011). Veri ambarı özel hazırlanmış verileri hızlı ve kolay bir şekilde elde etmeyi amaçlar. Şirketlerin çeşitli lokasyonlarda bulunan veri depolarından alınan bu özel veriler, yönetim raporlamalarında, çeşitli işyeri sorgularında, karar destek sistemlerinde, yönetici bilgi sistemlerinde ve veri madenciliği uygulamalarında kullanılır (Ponniyah, 2011).

Veri ambarları; doksanlı yılların başlarından beri, karar destek sistemlerinin başarısında önemli rol oynar. Kuşkusuz, veri ambarı kullanıcıları verilerin doğru ve

eksiksiz olduğundan emin olma ihtiyacı duyar. Veri ambarının, geçici olmayan ve zamanla değişen verileri içermeye özelliği, hali hazırda var olan verileri yok etmeden veriler üzerinde değişiklik yapmaya izin vermeyi gerektirir.

OLAP (online analytical processing) çevrimiçi analitik işleme canlı, gerçek ve hazır dataya ulaşma ihtiyacından dolayı ortaya çıkmıştır. OLAP raporları ile istatistiksel güvenilirliğe sahip gelecek durum analizleri yapmak mümkündür. OLAP küplerindeki veriler günün belli zamanlarında güncellenir ve işlenir. Toplam, ortalama hesaplama gibi tüm hesap işlemleri veri değiştiğinde yeniden hesaplanır. Rapor OLAP küpleri ile sunulduğunda arka planda hiçbir hesaplama yapılmaz, bir öndeki veri güncellemesinde hesaplanmış veriler gösterilir. Tüm hesaplanan değerler daha önce olduğu gibi OLAP küplerinde saklanır. Raporlama esnasındaki yegâne işlem, raporu çağırmak ve kullanıcıya göstermektir. Subotić ve arkadaşları (2014) ve Ravat ve Teste (2006) OLAP boyutları üzerinde versiyonlama ve dinamik güncelleme çalışmalarda yapılmıştır. Tipik OLAP işlemleri arasında bir ya da birden çok boyut hiyerarşisi arasında veri toplama seviyesini yükseltme (rollup) ve veri toplama seviyesini azaltma (drill-down), küçük bir parçayı seçip gösterme (slice_and_dice), ve verilerin çok boyutlu görünümünün yeniden yönlendirilmesi (pivot) bulunmaktadır (Ceci ve Ark,2013; Wrembel, 2009).

Zamansal veri tabanı çalışmaları bu alanda önemli sonuçlar ortaya çıkarmıştır. Tarihsel bir veri tabanı ilgili nesne ya da özelliğin geçmiş zamandaki bilgisini tutabileceği gibi (olay zamanı), veri tabanındaki işlem geçmişini de (işlem zamanı) tutabilir. Olay zamanı nesnelerin ya da özelliklerin zamanla değişen durumlarını modellemek için kullanılır. Bu durumlar geçmişte, günümüzde ya da gelecekte olabilir. İşlem zamanı nesne ya da özelliğin veri tabanına kaydolması ve sonrasında yapılan her türlü değişikliğin zaman etiketini belirtir. Günümüzde kullanılan veri tabanlarında işlem zamanı şimdiki zamanın ötesinde

olamaz, geçmiş zaman ya da gelecek zaman bilgilerini içermez. Hâlbuki çift zamanlı veri tabanları değişen dünyanın değişen bilgilerini modelleyebilir, verilerden elde edilmiş gerçeklerin olay zamanlarını ve nesnelere üzerinde yapılan veri tabanı işlemlerinin geçmişini tutabilir (Atay ve Tansel, 2009).

Veri ambarlarında zamansal bilgi depolanır, bununla ilişkili olarak zamansal veri tabanları üzerinde yapılan araştırmalar, veri ambarları açısından da yararlıdır. Zamansal veri ambarları, zamana bağlı değişen çok boyutlu verilerin yönetimini sağlar. Daha önceden de zamansal veri tabanları ve veri ambarlarını bir araya getiren araştırma projeleri yapılmıştır. Malinowski ve Zimanyi (2006) yılında yaptıkları çalışmada, zamana bağlı olarak değişen ve zamana bağlı olarak değişmeyen çok boyutlu bir model sunmuşlardır. Martin ve Abello (2003) veri ambarları ve zamansal veri tabanları arasındaki benzerlik ve farklılıkları defatlı bir şekilde yaptıkları çalışmada sunmuşlardır. Eder ve arkadaşları (2001) çalışmalarında, zamansal veri ambarı modelini iki farklı yaklaşım kullanarak sunmuşlardır. Bunlardan biri olan doğrudan yaklaşımda, her sonucunu elde etmek için bir sorgu analizörü ve sonuç analizörü kullanılır. Bir diğeri doğrudan olmayan yaklaşımda, data mart adı verilen küçük veri ambarları belli yapısal özelliklere göre oluşturulur. Diğer tüm yapı versiyonlarındaki veriler gerekli dönüşüm fonksiyonları ile işlenerek yeni oluşturulan data marta aktarılır. Doğrudan yaklaşım daha fazla esneklik sunarken, dolaylı yaklaşım data mart bir kere oluşturulduktan sonra tepki süresi açısından üstündür. Janet ve arkadaşları (2006) veri ambarı hali hazırda var olan veri ambarı küpleri üzerindeki boyutların değiştirilmesi, silinmesi, küplere yeni boyut eklenmesi gibi işlemleri çalışan uygulamalara zarar vermeden yapan bir model geliştirmişlerdir. Ayrıca çok boyutlu şemalarda; ekleme, silme, değiştirme işlemlerini yapabilmek için bir de sorgulama dili sunmaktadırlar. Combi ve arkadaşları (2009) zamansal ve yarı yapılı veri ambarları için grafik tabanlı bir model ve gerekli bilgiyi kullanışlı bir şekilde çekmeyi sağlayan

bir sorgulama dili sunmuştur.

Bu çalışmada; veri ambarı ve OLAP uygulamalarının öznelik üzerinde olay ve işlem zamanı bilgilerini içeren, çift zamanlı bir uzantısı yapılmıştır. Veri ambarı ve zamansal veri tabanları alanlarını birbirlerine bağlayan araştırma projeleri daha önceden yapılmıştır. Ancak, şu ana kadar yapılan çalışmalarda zaman etiketleri satır üzerindedir. Bizim bilgimize göre veri ambarı ve öznelik zaman etiketi daha önce birlikte kullanılmamıştır.

Makalenin geri kalanı şu şekilde organize edilmiştir. İkinci bölüm, veri ambarları hakkında özet bilgi içerir. Üçüncü bölümde, çift zamanlı veri tabanları ve iki zamanlı etiketleme yaklaşımları ana hatları ile anlatılmıştır. Dördüncü bölümde, çift zamanlı veri ambarları için sunulan model ve uygulama detayları sunulmuştur. Beşinci bölümde, deneysel çalışma için kullanılan örnek sorgular ve test sonuçları açıklanmıştır. Altıncı bölümde çalışmadan çıkan sonuç özetlenmiş ve gelecek çalışmalarda yapılması planlanan işlere değinilmiştir.

Veri ambarları

Veri ambarı şemasında üç temel kavram kullanılır; olaylar (fact), boyutlar (dimension) ve ölçekler (measure). Bir olay, ölçeklerden ve bağlam verilerinden oluşan ilişkili veri öğelerinin bir derlemidir. Her olay tipik olarak bir iş öğesi, iş hareketi ya da iş sürecini çözümlenmekte kullanılabilecek olan bir oluşu ifade eder. Bir veri ambarında, olaylar sayısal verinin saklandığı ana tablolarda gerçekleştirilmektedir. Bir boyut, aynı türdeki görüntülere ait üye ya da birimlerin bir derlemidir. Olay tablosundaki her veri noktası, çok sayıdaki boyutların her birinden bir ve yalnızca bir üyeye ilişkilendirilir. Boyutlar olaylar için kavramsal altyapıyı sağlarlar. Birçok çözümlenme işlemi olaylar üzerindeki boyutların etkisini sayıya dökmek üzere kullanılmaktadır. Boyutlar, üzerinde Çevrimiçi Çözümlenme İşlemi (Online Analytical Processing-

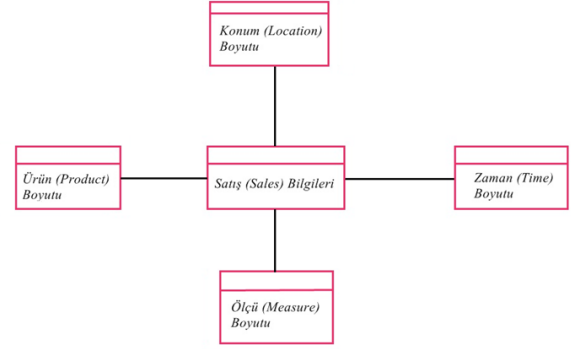
OLAP) yapılacak olan parametrelerdir. Bir boyutta birçok boyut üyesi vardır ve bunlar bir veri ögesinin yerini belirlemek üzere kullanılan ayrı bir isim ya da belirleyicidir. Örneğin, bütün şehirler, bölgeler ya da ülkeler de bir coğrafya boyutu oluştururken, bütün aylar ya da yıllar bir zaman boyutu oluşturmaktadır. Ölçek bir olayın sayısal niteliğidir. Bu nitelik, işin boyutlara göreli olarak başarımını gösterir. Sayılar değişken olarak adlandırılır. Örneğin ürün satışlarının parasal miktarı, satış hacmi, hareket miktarı gibi değerler ölçek olarak düşünülebilir. Ölçekler, boyut üyelerinin birleşimleriyle belirlenir ve olaylar üzerine yerleştirilir.

Veri ambarı şemasında kullanılacak “Yıldız modeli” ve “Kar tanesi modeli” olarak iki temel model bulunmaktadır: Yıldız modeli veri ambarı için temel yapıdır. Genellikle bir büyük merkezi tablo (olay tablosu olarak adlandırılır) ve olay tablosu etrafında yıldız biçiminde sıralanmış daha küçük tablolar söz konusudur. Şekil 1’de; merkezde bir satış bilgileri tablosu ve satış bilgileri tablosu etrafında sıralanmış zaman, ürün, konum ve ölçü boyut tablolarını gösteren, bir yıldız modeli örneği görülmektedir.

Kar tanesi modeli bazen kendi sıra düzenlerini de içeren bir ya da daha fazla boyutun ayrıştırılmasının sonucu olarak ortaya çıkar. Bir boyut tablosunun üyeleri arasındaki birçok-bir ilişkileri, bir sıra düzen oluşturacak şekilde ayrı bir boyut tablosu olarak tanımlanabilir. Örneğin, Şekil 1’deki ürün boyutu, Şekil 2’de model, kategori ve alt-kategori boyutlarına ayrıştırılmıştır. Yıldız modelinden türetilen bu model, görüldüğü gibi bir kar tanesine benzemektedir.

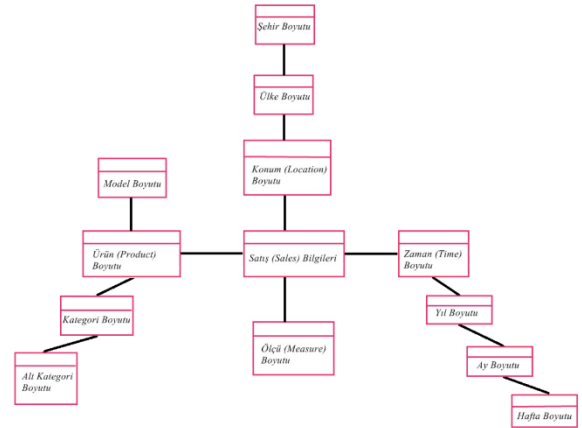
Bir boyutlu modeli görselleştirmenin en yaygın yolu bir “küp” (cube) çizmektir. Üç boyutlu bir model küp kullanılarak gösterilebilir. Boyutlu model genellikle üçten fazla boyut içerdiğinden hypercube olarak adlandırılır. Ancak bir hypercube’ün görselleştirilmesi

zor olduğundan, genel olarak kullanılan terim küptür.



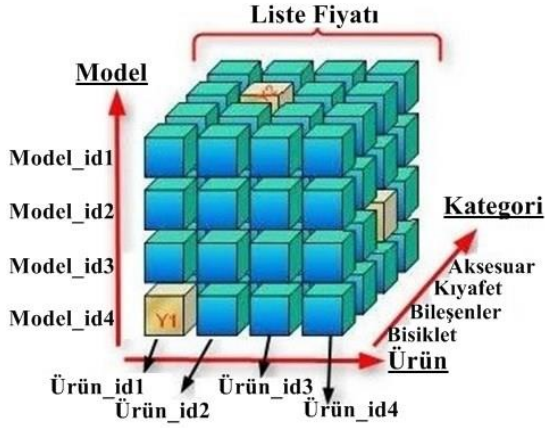
Şekil 1. Yıldız modeli.

Bu kavramları, ürünün liste fiyatını kategorisine ve modeline göre yapılan analizler ile açıklayacağız. Her bir ürünün liste fiyatı, sayı ile nitelendirilebilen bir gerçektir. Alt kategori ve model ürün boyutunun ölçüleridir. Her bir ürün tipi için farklı bir sayısal ölçü değeri vardır. Bu tip bağlantı boyutlara göre gerçeklerin organize olması ile oluşur.



Şekil 2. Kar Tanesi modeli.

Gerçek durum analizinde ortaya çıkabilecek sonuçlar boyut ve ölçü arasındaki ilişkiye ve veri ambarı şemasına bağlıdır. İhtiyaçlara uygun boyut, ölçü gibi veri ambarı nesnelere saptanır. Şekil 3 de liste fiyatının ürün, model ve alt kategori boyutlarına göre analiz edildiği çok boyutlu bir küp sunulmuştur.



Şekil 3. Çok boyutlu modelin küp görünümü sunumu.

Çift zamanlı veri tabanları

Tarihsel veri tabanları veri tabanı durumlarını nesnelerin olay zamanı ekseninde saklar, ancak nesnelerin sistemde yaşadıkları değişiklikleri içermez. İşlem veri tabanları sistem değişikliklerini saklar, ama nesnelere hakkında zamansal bilgi içermez. Nesnelerin zaman içindeki değişimleri hakkında tam bilgi sahibi olabilmek için tarihsel ve işlemsel veri tabanlarının bir arada kullanılması gerekmektedir. Çift zamanlı veri tabanı bu gereksinimleri karşılar. Böylelikle, çift zamanlı veri tabanları, gerçek verilerinin değişimini zaman üzerinde doğru bir şekilde ifade etmeyi sağlar (Ben-Zvi, 1982; Golfarelli ve Rizzi, 2011).

Zamansal veri tabanlarının uygulanmasında çok bilinen iki yaklaşım vardır. Bunlardan biri satır zaman etiketlemedir ve birinci normal form ilişkilerini kullanır (1NF). Bir diğeri öznitelik zaman etiketidir ve normal form ilişkilerini içermesi gerekmez. Satır zaman etiketleme yaklaşımında zamanın satır üzerinde belirtildiği durumlarda özellikler aynı anda değişmiyorsa, her bir özellik için ayrı bir tablo oluşturmak gerekir. Zamanla değişmeyen özellikler ise başka bir ilişki ile gruplandırılır. Satırlara zaman etiketi eklemek için, dört tarihsel özellik satıra eklenir. Satır zaman etiketleme, geleneksel ilişkisel veri tabanlarının tüm avantajlarından faydalanır. Ancak zamanla değişen

her özellik için ayrı bir zaman bilgisi vardır ve ayrı bir ilişki ile gösterilmelidir. Örneğin, ürünün liste fiyatının geçerlilik aralığı, olay başlangıç zamanı (VT_LB) ve olay bitiş zamanı (VT_UB) arasında geçen süredir. Eğer bir kayıt şu an geçerli ise VT_UB değeri 'şimdi' (now) ye eşittir. Buna ek olarak, geçmiş ve şu anda gerçekleşen veri tabanı işlemlerinin ilk işlem zamanı (TT_LB) ve son işlem zamanı (TT_UB) bilgileri de tutulur. Tablo 1'de görülen MB1 ürününün eski liste fiyatı 1191.17'dir ve ilk işlem zamanı 24.09.2002'dir. MB1 ürününün liste fiyatı 01.07.2002 tarihinde değişip 1226.9 olduğunda tabloya yeni bir satır eklenmiştir. Eski liste fiyatı bilgisinin olay bitiş tarihi ve son işlem tarihi 01.07.2002 olarak değişmiştir. Yeni liste fiyatı bilgisinin ise olay başlangıç tarihi ve ilk işlem tarihi bilgileri 02.07.2002 olarak kaydedilmiştir, yeni bilgi geçerli olan bilgi olduğundan VT_LB değeri şimdiyi (now) göstermektedir.

Öznitelik zaman etiketleme durumunda normal form özellikleri kullanılmaz ve zaman etiketleri özniteliklere bağlıdır. Modellenen nesne ile ilgili tüm bilgiler tek bir satırda saklanabilir. Nesneye ait herhangi bir değer değiştiğinde, yenilenmeden etkilenmeyen diğer değerlerin (Ürün adı, ürün rengi gibi) tekrarlanması gerekmez. Zamanla değişen ve değişmeyen tüm özellikler bir bütün olarak tek bir tablo içinde modellenebilir.

Çift zamanlı atom (<işlem zamanı, olay zamanı, değer>) olarak gösterilen üçlüdür. <[VTl, VTu), [TTl, TTu), V> formundaki çift zamanlı atomda, VT_LB olay başlangıç zamanını, VT_UB olay bitiş zamanını, TT_LB işlem başlangıç zamanını, TT_UB işlem bitiş zamanını ve V nesne değeri olarak ifade edilir. Tablo 2' deki çift zamanlı atom

{<[25.09.2001, 01.07.2002], [24.09.2001, 01.07.2002], 1191.17>,

<[02.07.2002, now], [02.07.2002, now], 1226.9>}

Tablo 1. Satır Zaman Etiketli Çift Zamanlı Yaklaşım.

P. Name	Color	List Price	VT_LB	VT_UB	TT_LB	TT_UB
MB1	Black	1191.17	25.09.2001	01.07.2002	24.09.2001	01.07.2002
MB1	Black	1226.9	02.07.2002	now	02.07.2002	now
MB2	Yellow	120.43	18.10.2002	now	20.10.2002	now
MP	Red	150	30.07.2002	now	01.08.2002	now
MA	Black	35.89	18.10.2002	now	20.10.2002	now

MB1 adlı ürünün veri tabanına 24.09.2001 tarihinde yazıldığını, efektif olay başlama zamanının 25.09.2001 olduğunu belirtir. Yaklaşık olarak on ay sonra, 01.07.2002 tarihinde yapılan bir işlemle ürünün fiyatı 1226.9'a yükseltilmiştir. Bir güncelleme işlemi yapıldığında, liste fiyatı özelliğine yeni bir satır eklenir, diğer özelliklerde değişim olmaz.

Çift zamanlı veri ambarı modeli ve uygulaması

Bu çalışmada kullanılan veriler araştırma ve deney amaçlı kullanılabilen Microsoft Adventure Works (AW) veri tabanından alınmıştır. AW veri tabanında beş farklı modül bulunmaktadır. Biz bu çalışmada kendi metodumuzu Production (Ürün) modülü üzerinde uyguladık. Ürün modülü şemasını öznitelik zaman etiketli ve normal formda olmayacak şekilde iç içe tablolar halinde yeniden oluşturduk.

AW yapısına uygun olarak, bazı tablolar gereksiz tekrarları önlemek için kendileri ile ilişkili tabloların içine gömüldü. AW Ürün modülünde 25 tablo varken, oluşturduğumuz sistemde bu bilgilerin tamamını içeren 15 tablo bulunmaktadır.

Nesnel ilişkili veri tabanı bize kullanıcı tarafından yeni tip (type) tanımlama imkanı sunar. Şekil 4'de çift zamanlı atom tipinin tanımlanmasını sağlayan veri tabanı sorgusu sunulmuştur. Tip bir kere tanımlandıktan sonra, veri tabanı içinde kendiliğinden gelen tipler gibi SQL sorgularında kullanılabilir.

Kullanıcı tanımlı tipler, tablonun tamamını oluşturan bir "veri tipi" olarak tanımlanabilir. Seçme, ekleme, silme gibi veri işlemleri ilişkisel veri tabanı

tablolarındaki gibi yapılabilir. İç içe bir tablo aynı tipteki elemanların oluşturduğu sırasız bir veri koleksiyonudur. Ayrıca iç içe tablolardaki özellik sayısı herhangi bir sınırlandırma yoktur.

```
CREATE TYPE BT_NUMBER AS (
    VT_LB DATE,
    VT_UB DATE,
    TT_LB DATE,
    TT_UB DATE,
    VALUE NUMBER);
```

Şekil 4. Kullanıcı tanımlı tip olarak çift zamanlı atom BT_NUMBER tanımlama.

AW veri tabanındaki tablo yapısı analiz edilip yeniden dizayn edilmiştir. Şekil 5'de LISTPRICE, PRODUCTSELL, ve PRODUCTCOST tiplerinin sütun zaman etiketli ve yarı yapıları oluşturulması için kullanılan sorgu bulunmaktadır. Bu çift zamanlı tablolar sadece BT_NUMBER'dan oluşan tek özellikli tablolarıdır.

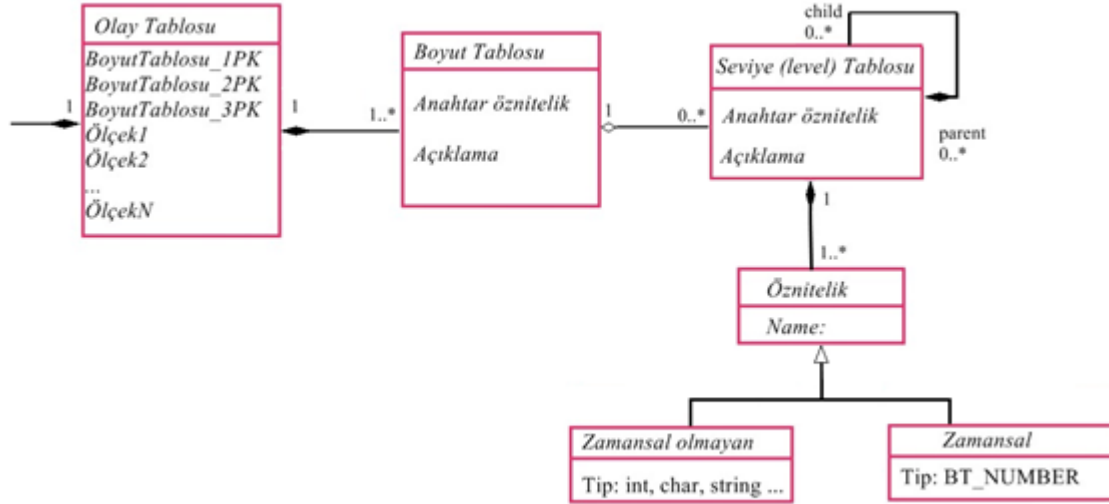
```
CREATE TYPE LISTPRICE AS TABLE OF
BT_NUMBER;
CREATE TYPE PRODUCTSELL AS TABLE OF
BT_NUMBER;
CREATE TYPE PRODUCTCOST AS TABLE OF
BT_NUMBER;
```

Şekil 5. Liste fiyatı (LISTPRICE) satırının BT_NUMBER tipinde iç içe tanımlanışı.

Kavramsal modeller, uygulamanın kullanıcı ihtiyaçlarına en yakın şekilde tanımlanmasını sağlar. Kavramsal modeller tasarımcı ve kullanıcı arasındaki iletişimin kolaylaşmasını sağlar [8].

Tablo 2. Öznitelik Zaman Etiketli Çift Zamanlı Yaklaşım.

P. Name	Color	List Price
MB1	Black	{<[25.09.2001, 01.07.2002], [24.09.2001, 01.07.2002], 1191.17 >, <[02.07.2002, now], [01.07.2002, now], 1226.9>}
MB2	Yellow	<[18.10.2002, now], [20.10.2002, now], 120.43>
MP	Red	<[30.07.2002, now], [01.08.2002, now], 150>
MA	Black	<[18.10.2002, now], [20.10.2002, now], 35.89>
TR	Black	<[30.07.2002, now], [01.08.2002, now], 22.11>

**Şekil 6.** Çok boyutlu çift zamanlı veri ambarının kavramsal modeli.

Önerilen çok boyutlu, yarı yapılı veri ambarı modeli Şekil 6' de sunulmuştur.

Bir olay tablosu üç veya daha fazla boyut tablosu ile ilişkilidir. Boyutlar bir veya birden çok seviyeden oluşur. Boyutların ve seviyelerin anahtar ve açıklayıcı alanları vardır. Özellikler adları int, real, varchar gibi veri tipleri ile ifade edilir. Seviyelerdeki özellikler için çift zamanlılık, çok özellikli olarak tanımlanan BT_NUMBER veri tipi ile sağlanır. En az bir zamansal seviyesi olan seviyeler zamansal seviyelerdir. En az bir zamansal seviye içeren boyutlar zamansal boyutlardır. Seviyeler ve boyutlar iç içe tablo olarak saklanan birçok zamanla değişen özellik barındırabilir.

AW veri tabanını analiz ettikten sonra, beş farklı küp dizayn edilmiştir.

Şekil 5'de LISTPRICE, PRODUCTSELL, ve PRODUCTCOST tiplerinin sütun zaman etiketli ve

yarı yapılı oluşturulması için kullanılan sorgu bulunmaktadır. Bu çift zamanlı tablolar sadece BT_NUMBER'dan oluşan tek özellikli tablolardır.

Kavramsal modeller, uygulamanın kullanıcı ihtiyaçlarına en yakın şekilde tanımlanmasını sağlar. Kavramsal modeller tasarımcı ve kullanıcı arasındaki iletişimin kolaylaşmasını sağlar [8]. Önerilen çok boyutlu, yarı yapılı veri ambarı modeli Şekil 6' de sunulmuştur.

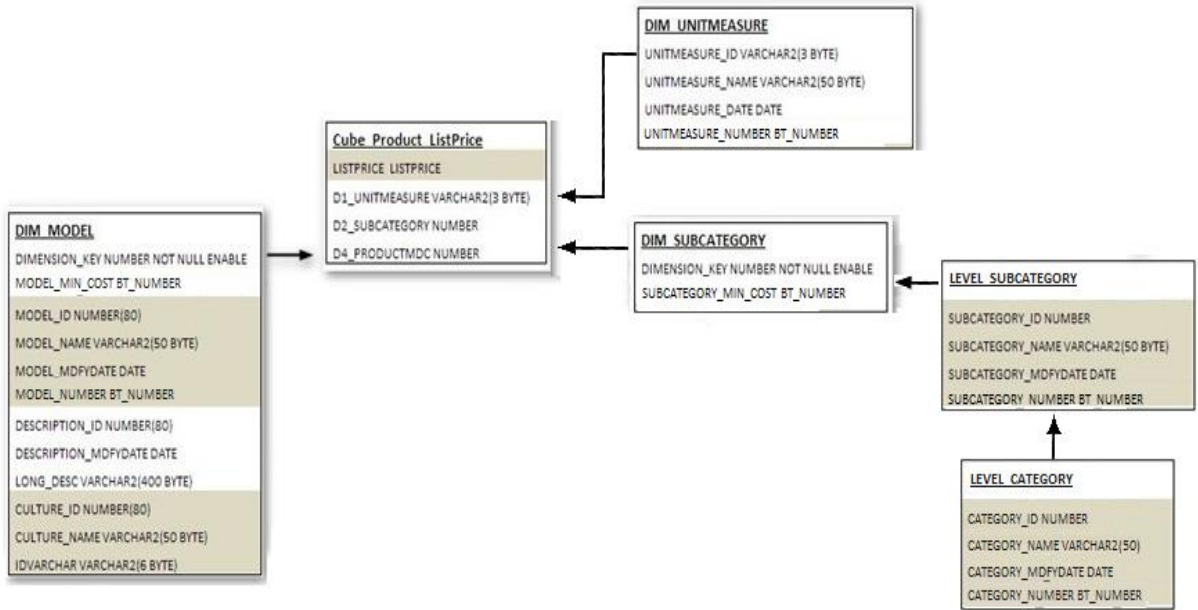
Bir olay tablosu üç veya daha fazla boyut tablosu ile ilişkilidir. Boyutlar bir veya birden çok seviyeden oluşur. Boyutların ve seviyelerin anahtar ve açıklayıcı alanları vardır. Özellikler adları int, real, varchar gibi veri tipleri ile ifade edilir. Seviyelerdeki özellikler için çift zamanlılık, çok özellikli olarak tanımlanan BT_NUMBER veri tipi ile sağlanır. En az bir zamansal seviyesi olan seviyeler zamansal seviyelerdir. En az bir zamansal seviye içeren boyutlar zamansal boyutlardır.

Seviyeler ve boyutlar iç içe tablo olarak saklanan bir çok zamanla değişen özellik barındırabilir.

AW veritabanını analiz ettikten sonra, beş farklı küp dizayn edilmiştir. Bunlar; Product, ProductListPrice, ProductCostHistory, ProductCost History2Dim, ProductListPrice2Dim ve Workorder_Order-Qty isimli küplerdir. Şekil 5 de gösterilen Cube_Product_ListPrice gerçek tablosundaki, D1_Unitmeasure, D2_Subcategory ve D3_Product MDC özellikleri boyut tabloları ile ilişki kurmaya yarayan özelliklerdir, BT_NUMBER tipindeki LISTPRICE özelliği ise ölçü değerini saklar.

Cube_Product_ListPrice küpü; Dim_Model, Dim_UnitMeasure ve Dim_Subcategory olmak üzere

üç boyuttan oluşmuştur. Dim_Model boyut tablosunda, Dimension_Key özelliği birincil anahtardır ayrıca BT_NUMBER veri tipinde Model_Min_Cost ve Model_Number adında iki özellik içerir. Dim_UnitMeasure tablosunda Unitmeasure_Id özelliği birincil anahtardır BT_NUMBER veri tipinde Unitmeasure_Number özelliği bulunur. Dim_Subcategory boyut tablosunda Dimension_Key özelliği birincil anahtardır, bunun yanında BT_NUMBER tipinde Subcategory_Min_Cost özelliği bulunur. Dim_Subcategory boyutunda iki seviye vardır. Şekil 7 de görüldüğü gibi Subcategory isimli seviyede BT_NUMBER veri tipinde Subcategory_Number adlı, Category isimli seviyede ise Category_Number adlı özellikler bulunur.



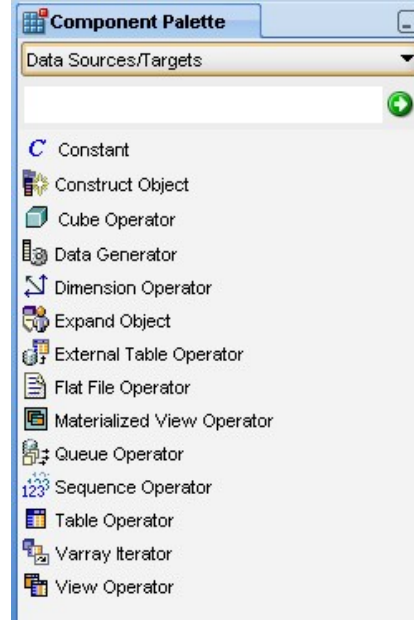
Şekil 7. Cube_Product_ListPrice tablosunu baz alan kar tanesi şeması.

Gerekli veri ambarı nesnelere Warehouse Builder aracı ile oluşturulmuştur. Veri ambarı düzenli bir şekilde oluşturulduktan sonra, sıra bu nesnelere saklanacak verileri, veri tabanından çekilen verilerle eşlemeye gelir. Warehouse Builder aracının içinde veri eşlemeyi sağlayan "Mapping" adı verilen bir araç vardır. Bu araç veri tabanındaki verileri alır ve hedef veri ambarı depolama alanına düzeni bozmayacak şekilde aktarır. Eşleme aracında veri ambarına aktarılacak verileri

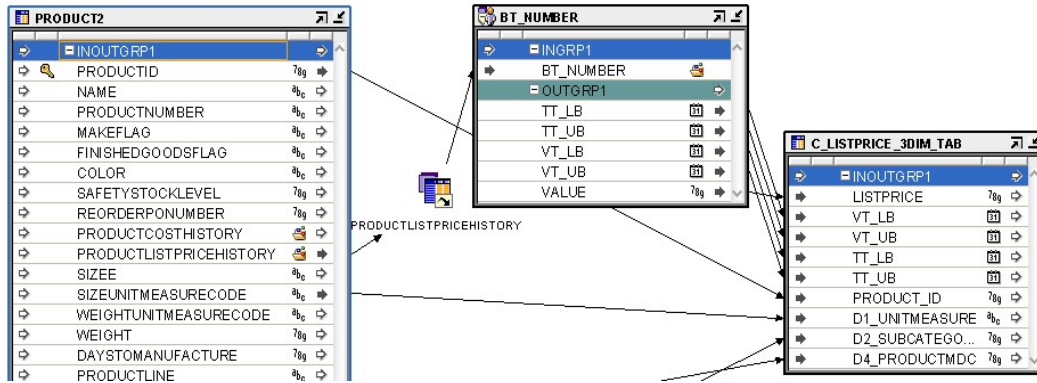
düzenlemek için, çok boyutlu nesnelere ve veri tabanı nesnelere sürüklenip bırakılabildiği bir tasarım arayüzü vardır. Verileri işlemek için gerekli her türlü bileşen, bileşen paletinde bulunur. Veri eşleme bileşen paleti Şekil 8 de gösterilmektedir.

Şekil 9 da örnek bir eşleme diyagramı gösterilmiştir. PRODUCT2 veri tabanı tablosu içindeki veriler, sağ taraftaki küp tablosuna eşlenmiştir. Veri ambarı nesnelere alanlar direkt olarak iç içe tipte

tanımlanamaz. Veri tabanında tanımlı iç içe bir tip olan BT_NUMBER tek bir alanda tutulabilirken, veri ambarında BT_NUMBER'ı oluşturan her bir özellik <VT_LB, VT_UB, TT_LB, TT_UB, VALUE> ancak ayrı birer alanda tutulabilir. Veri ambarı altyapısı henüz ne yazık ki iç içe tipleri kabullenmeye hazır değildir. Bu sebeple veritabanındaki tablolarında iç içe tipteki alanların, veri ambarı nesnelere ile eşlenirken, bazı işlemlerden geçmesi gerekir. Eşleme ara yüzünde bulunan genişleme nesnesi (Expand Object), iç içe tipleri giriş, iç içe tiplerin içindeki her bir alanı ise çıkış değeri olarak düzenler.



Şekil 8. Veri işleme bileşen paleti.



Şekil 9. Örnek bir eşleme tasarım sayfası.

Çift zamanlı veri ambarı sorguları

Sunulan modelin uygulanabilirliğini göstermek adına; aynı sonucu veren sorgular, hem ilişkisel veritabanı nesnelere kullanarak, hem de çok boyutlu çift zamanlı veri ambarı nesnelere kullanarak çalıştırılmıştır. Sorgular belirli zaman aralıklarında altışar kez çalıştırılmış ve bu altı çalıştırmanın ortalaması alınmıştır. Şekil 10 da üç farklı sorgu için, ortalama sorgulama zamanları listelenmiştir. Sorgu 3 ün diğer sorgulardan daha uzun sürece çalıştığı açık bir şekilde gözlemlenmektedir. Bunun sebebi, Sorgu 3 ün veri çektiği tablonun kayıt sayısının fazla olmasıdır.

Sorgu 1: Kategorilere, alt kategorilere ve yıllara göre ortalama bisiklet fiyatları nedir?

Aynı sonucu veren sorgular Tablo 3'te gösterildiği gibi iki farklı ortamda yazılmıştır. Veri ambarı sorgusu sonucu elde edebilmek için CUBEPRODUCTLISTPRICE küp tablosundan veri almaktadır, ayrıca her biri farklı bir sınırlandırma içinde olmak üzere iki tane iç içe sorgu içerir. İlişkisel veritabanı sorgusu aynı sonucu elde edebilmek için, dört farklı tablodan veri çeker ve bu tabloları birbiri ile ilişkilendirebilmek için üç farklı sınırlandırma koşulu içerir.

Sorgu 2: Modellere göre ortalama bisiklet fiyatları nedir?

Tablo 3. Sorgu 1, Veri ambarı ve ilişkişel veri tabanı nesnelere ile

Veri Ambarı:
SELECT VT_LB,VT_UB,TT_LB,TT_UB,ROUND(AVG(LISTPRICE),2), (SELECT CATEGORY_NAME FROM DIM_SUBCATEGORY WHERE DIMENSION_KEY= DIM_SUBCATEGORY_KEY) CATEGORY_NAME, (SELECT SUBCATEGORY_NAME FROM DIM_SUBCATEGORY WHERE DIMENSION_KEY= DIM_SUBCATEGORY_KEY) SUBCATEGORY_NAME FROM CUBEPRODUCTLISTPRICE GROUP BY VT_LB,VT_UB,TT_LB,TT_UB,DIM_SUBCATEGORY_KEY ORDER BY DIM_SUBCATEGORY_KEY;
İlişkişel Veri Tabanı:
SELECT STARTDATE AS VT_LB, ENDDATE AS VT_UL, PLP.MODIFIEDDATE AS TTLB, PLP.MODIFIEDDATE AS TT_UB, PC.NAME AS CATEGORY_NAME, PSC.NAME AS SUBCATEGORY_NAME, AVG(PLP.LISTPRICE) FROM PRODUCTION.PRODUCTLISTPRICEHISTORY PLP, PRODUCTION.PRODUCTSUBCATEGORY PSC, PRODUCTION.PRODUCT P, PRODUCTION.PRODUCTCATEGORY PC WHERE PLP.PRODUCTID=P.PRODUCTID AND PSC.PRODUCTSUBCATEGORYID=P.PRODUCTSUBCATEGORYID AND PC.PRODUCTCATEGORYID=PSC.PRODUCTCATEGORYID GROUP BY STARTDATE,ENDDATE,PLP.MODIFIEDDATE,PC.NAME,PSC.NAME;

Veri ambarı sorgusu sonucu elde edebilmek için CUBEPRODUCTLISTPRICE küp tablosundan veri almaktadır, ayrıca bir tane iç içe sorgu içerir. İlişkişel veri tabanı sorgusu aynı sonucu elde edebilmek için, üç farklı tablodan veri çeker ve bu tabloları birbiri ile ilişkilendirebilmek için iki farklı sınırlandırma koşulu

çerir. Veri ambarı sorgusu bir tablodan veri çekerken, ilişkişel veri tabanı sorgusu üç farklı tablodan verileri çekmek için daha fazla bağlantı ve arka planda daha fazla işlem yapmaktadır. Aynı sonucu veren sorgular Tablo 4'te gösterilmektedir.

Tablo 4. Sorgu 2, Veri ambarı ve ilişkişel veri tabanı nesnelere ile

Veri Ambarı:
SELECT ROUND(AVG(LISTPRICE),2), (SELECT MODEL_NAME FROM DIM_MODEL WHERE DIMENSION_KEY=DIM_PRODUCTMDC_KEY) MODEL_NAME FROM CUBEPRODUCTLISTPRICE GROUP BY DIM_PRODUCTMDC_KEY ORDER BY DIM_PRODUCTMDC_KEY;
İlişkişel Veri Tabanı:
SELECT STARTDATE AS VT_LB, ENDDATE AS VT_UL, PLP.MODIFIEDDATE AS TTLB,PLP.MODIFIEDDATE AS TT_UB, PM.NAME AS MODEL_NAME,AVG(PLP.LISTPRICE) FROM PRODUCTION.PRODUCTLISTPRICEHISTORY PLP, PRODUCTION.PRODUCTMODEL PM, PRODUCTION.PRODUCT P WHERE PLP.PRODUCTID=P.PRODUCTID AND PM.PRODUCTMODELID=P.PRODUCTMODELID GROUP BY STARTDATE,ENDDATE,PLP.MODIFIEDDATE,PM.NAME;

Sorgu 3: İş emirlerini, kategori ve alt kategoriye göre artan sıra ile listeleyiniz.

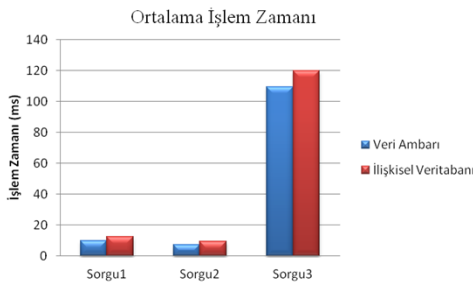
Veri ambarı sorgusu sonucu elde edebilmek için CUBE_WORKORDER_ORDERQTY küp tablosundan ve DIM_PRODUCT boyut tablosundan

veri almaktadır. Bu tabloları ilişkilendirmek için bir tane sınırlandırma koşulu içerir. İlişkişel veri tabanı sorgusu aynı sonucu elde edebilmek için, dört farklı tablodan veri çeker ve bu tabloları birbiri ile ilişkilendirebilmek için üç farklı sınırlandırma koşulu içerir. Yazılan sorgular Tablo 5'ta gösterilmektedir.

Tablo 5. Sorgu 3, Veri Ambarı ve İlişkisel Veritabanı Nesnelere ile

<p>Veri Ambarı:</p> <pre>SELECT MAX(CUBE_WORKORDER_ORDERQTY.VALUE) ORDERQTY, DIM_PRODUCT.CATEGORY_NAME, DIM_PRODUCT.SUBCATEGORY_NAME, DIM_PRODUCT.PRODUCT_NAME FROM CUBE_WORKORDER_ORDERQTY, DIM_PRODUCT WHERE DIM_PRODUCT.DIMENSION_KEY= CUBE_WORKORDER_ORDERQTY.DIM_PRODUCT_KEY GROUP BY DIM_PRODUCT.CATEGORY_NAME, DIM_PRODUCT.SUBCATEGORY_NAME, DIM_PRODUCT.PRODUCT_NAME ORDER BY DIM_PRODUCT.CATEGORY_NAME , DIM_PRODUCT.SUBCATEGORY_NAME, DIM_PRODUCT.PRODUCT_NAME</pre>
<p>İlişkisel Veri Tabanı:</p> <pre>SELECT MAX(WO.ORDERQTY) ORDER_QTY, PS.NAME SUBCATEGORY_NAME, PC.NAME CATEGORY_NAME, P.NAME PRODUCT_NAME FROM PRODUCTION.WORKORDER WO, PRODUCTION.PRODUCT P, PRODUCTION.PRODUCTSUBCATEGORY PS, PRODUCTION.PRODUCTCATEGORY PC WHERE P.PRODUCTID=WO.PRODUCTID AND P.PRODUCTSUBCATEGORYID=PS.PRODUCTSUBCATEGORYID AND PS.PRODUCTCATEGORYID=PC.PRODUCTCATEGORYID GROUP BY PS.NAME, PC.NAME, P.NAME ORDER BY ORDER_QTY DESC</pre>

Çok boyutlu çift zamanlı model kullanıcı tanımlı tiplerle ve iç içe bağlantılarla tasarlanmıştır; bu nedenle veri ambarında sorgular aynı işi daha az tablo kullanarak yapabilir. Diğer taraftan ilişkisel veri tabanı sorgularını işletmek için çok sayıda tablodan veri çekilir ve bu tabloları birbiri ile ilişkilendirmek için anahtar ilişkiler belirlenir. Sonuç olarak, yapılan ölçümlerde ilişkisel veri tabanı sorgularının aynı işi yapabilmek için daha çok zaman harcadıkları tespit edilmiş ve Şekil 10 da elde edilen değerler sunulmuştur. Ayrıca ilginç bir şekilde, çok boyutlu modelin oluşturduğu tablolar segmentlerde 11,392KB yer kaplarken, ilişkisel veri tabanı tabloları 34.688KB yer kaplamaktadır. Bu ikinci yaklaşımın üç kat daha fazla bellek alanı gerektirdiğini ifade eder ki bu oldukça büyük bir rakamdır.

**Şekil 10.** Sorgular için yürütme zamanları.

Sonuç

Veri ambarı; müşteri ilişkileri yönetimi sistemleri (CRM-customer relationship management), küresel e-ticaret kuruluşları, tedarik zinciri yönetimi (SCM-supply chain management), kurumsal kaynak planlama sistemleri (ERP-enterprise resource planning), kurumsal bilgi portalları, strateji yönetimi ile yakından ilişkilidir. Gün geçtikçe veri ambarı adı geçen sistemlerin ayrılmaz bir parçası haline gelmektedir.

Bu çalışma, olay ve işlem zamanlarını öznelik zaman etiketleme yöntemiyle veri ambarına uygulanmasını açıklar ve deneysel örnek modelin uygulanabileceğini gösterir. Model hem olay hem de işlem zamanlarının veri ambarlarında kullanılacağını göstermektedir. Ayrıca, değişen dünyamızın değişen bilgilerini modelleyerek veri değerlerini eylemler ile birleştirir ve eylemlerin ne zaman geçerli olduğunu belirtir. Dolayısıyla veri değerlerinin geçmişini ve değişimlerinin bütün dökümünü sağlar. Gerçek dünyayı eksiksiz ve düzgün modelleyebilmek için her iki zaman boyutu da gereklidir.

Önerilen modelin uygulama performansı hakkında fikir edinmek için farklı sorgular çalıştırdık. Sonuçlar

sunulan modelin yalın veri tabanı uygulamasından daha iyi sonuç ve daha iyi performans ortaya koyduğunu gösteriyor.

Birçok firma daha etkili ve hızlı raporlama için veri ambarı teknolojilerini kullanmaktadır. Öznitelik zaman etiketli çift zamanlı yaklaşım daha az birleştirme deyimi kullanır ve arka planda daha az işlem yapar. Bu nedenle performansı daha iyidir. Yaptığımız çalışmanın veri ambarı çalışmalarına zemin hazırlamasını umut etmekteyiz.

Kaynaklar

ABELLÓ A., MARTÍN C. The Data Warehouse: An Object-Oriented Temporal Database. JISBD, 675-684, 2003.

ATAY C. E., TANSEL A. U. Bitemporal databases: Modeling and implementation, VDM Publishing, Germany, 2009.

BEN-ZVI J. The time relational model, Doktora Tezi, University of California, Los Angeles, 1982.

CECI M., CUZZOCREA A., MALERBA D. Effectively and efficiently supporting roll-up and drill-down OLAP operations over continuous dimensions via hierarchical clustering. Journal of Intelligent Information Systems, 1–25, 2013.

COMBI C., OLIBONI B., POZZI G. Modeling and querying temporal semistructured data warehouses. New Trends in Data Warehousing and Data Analysis Annals of Information Systems, 1-25, 2009.

EDER J., KONCILIA C., MORZY T. A. model for a temporal data warehouse. OES-SEO Workshop, 48-54, 2001.

GOLFARELLI M., RIZZI S. Temporal data warehousing: Approaches and techniques, PA: Information Science Reference, 1-18, 2011.

INMON W., KRISHNAN K. Building the Unstructured Data Warehouse, Technics Publications, New Jersey, 2011.

Bundan sonraki çalışmalarda, veri ambarı farklı zaman aralıklarında çift zamanlı yapının esnekliğinden faydalanılarak analiz edilebilir. Müşterilerin satın alma alışkanlıkları zamana bağlı analiz edilebilir. Yalın verimli dönemleri saptanabilir. Bu ve benzer yararlı durumlar; veri ambarı, kullanıcı tanımlı veri tipleri ve iç içe veri saklayabilmenin faydasıyla, diğer sistemlerden daha hızlı analiz edilebilir. Sonuçlar şirketlere stratejilerini belirlemede yardımcı olacaktır.

JANET E., RAMÍREZ R., GUERRERO E. B. A model and language for bitemporal schema versioning in data warehouses, IEEE, CIC '06. 15th International Conference, 309–314, 2006.

KAUFMANN M., FISCHER P. M., MAY N., TONDER A., KOSSMANN D. Tpc-bih: A Benchmark for Bitemporal Databases. Springer International Publishing, 8391, 16-31, 2014.

MALINOWSKI E., ZIMANYI E. A conceptual solution for representing time in data warehouse dimensions. 3rd Asia-Pacific Conference on Conceptual Modeling, 45–54, 2006.

PONNIAH P. Data warehousing fundamentals for IT professionals, John Wiley & Sons, 2011.

PORTER L. M. L. Bitemporal relational databases and methods of manufacturing and use. U.S. Patent Application, 614-609, 2012.

RAVAT F., TESTE O. Supporting Data Changes in Multidimensional Data Warehouses. International Review on Computers and Software, 251-259, 2006.

SUBOTIĆ D., POŠČIĆ P., JOVANOVIĆ V. Data Warehouse Schema Evolution: State of the Art, Centra, European Conference on Information and Intelligent Systems, 18-23, 2014.

WREMBEL R. A survey on managing the evolution of data warehouses. International Journal of Database Management Systems, IJDM, 2010.