Sigma J Eng Nat Sci, Vol. 42, No. 2, pp. 321−334, April, 2024

Sigma Journal of Engineering and Natural Sciences
Web page info: https://sigma.yildiz.edu.tr
DOI: 10.14744/sigma.2023.00076

sigma
Journal of Engineering and Natural Sciences
Mühendislik ve Fen Bilimleri Dergisi

Research Article

Comparison of feature-based sentence ranking methods for extractive summarization of turkish news texts

Ertürk ERDAĞI[1,*], Volkan TUNALI[2]

[1]*Department of Computer Engineering, Graduate School of Education, Maltepe University, Istanbul, 34857, Türkiye*
[2]*Department of Software Engineering, Faculty of Engineering and Natural Sciences, Maltepe University, 34857 Istanbul, Türkiye*

ARTICLE INFO

Article history
Received: 27 April 2022
Revised: 24 July 2022
Accepted: 09 November 2022

Keywords:
Summarization; Extractive; Sentence Ranking; Major Vowel; Minor Vowel

ABSTRACT

Document summarization is the task of generating a shorter form of document with important information content. Automatic text summarization has been developed for this process and is still widely used. It is divided into two main parts as extractive summarization and abstractive summarization. In this study, we used sentence ranking methods for extractive summarization for Turkish news text within the scope of the experimental study. We used different summarization rates, 20%, 30%, 40%, 50% and 60%. Summarization results were evaluated with the ROUGE ve BLEU metrics. We proposed new methods based on major vowel harmony and minor vowel harmony features. We obtained high evaluation results in both ROUGE ve BLEU metrics with major vowel harmony and minor vowel harmony features. Additionally, we studied a hybrid model using major vowel harmony and minor vowel harmony rules together. We obtained the best results with major vowel harmony, minor vowel harmony, and hybrid model (major vowel harmony and minor vowel harmony together). We compared the three proposed methods with the BERTurk model prepared for Turkish based on Google BERT. The results obtained gave very close results to this state-of-the-art method and showed that it is worth developing.

Cite this article as: Erdağı E, Tunalı V. Comparison of feature-based sentence ranking methods for extractive summarization of turkish news texts. Sigma J Eng Nat Sci 2024;42(2):321−334.

INTRODUCTION

The use of digital media in every aspect of daily life causes a large increase in the amount of data. Although this increase may seem to be a disadvantage, the increase in hardware and software facilities required for processing and interpreting data in recent years has provided various opportunities. The increase in the use of text-based data in different sectors and purposes as well as the increase in the amount of this data cause the reader to spend more time on important issues such as accessing summary information and extracting information.

Automatic text summarization systems have been developed to save time and quickly obtain the desired information from large-scale data. Summarization studies in text-based data are divided into two as single documents and multiple documents according to the number of sources. In terms of

***Corresponding author.**
*E-mail address: erturkerdagi@gmail.com

This paper was recommended for publication in revised form by Regional Editor Ahmet Selim Dalkilic

Published by Yıldız Technical University Press, İstanbul, Turkey

structure and semantics, studies are divided into two main parts as abstractive summarization and extractive summarization. Abstractive summarization proceeds in the form of reading the entire text and expressing it with new sentences and words partially independent of the main text [1]. In extractive summarization, it is determined whether the sentences will be included in the summary with various statistical and heuristic methods [2]. Abstractive summarization is used as a phenomenon close to human experience, it also brings along difficulties of grammatical and semantic structure in the studied language. The extractive summarization system follows a method about determination of the order of importance with a ranking made using attributes on the sentence and paragraph and including them in the summary based on this importance. In the ranking process using statistical methods, such examples can be given as assuming that the words that occur frequently in the text are important and that the first or last sentence is the one that reflects the text [3].

In this study, we examined the sentence structure in Turkish news texts and used various statistical features for the extractive summary. Although automatic text summarization is a frequently studied area in the literature, there are not many studies on Turkish texts [4]. We recommend the Major Vowel and Minor Vowel Harmony rules, which are important for the morphological rules of Turkish. We observed that these rules were more effective on extractive summarization, so we proposed a hybrid model and these two rules. Two rules are used in Turkish to check whether a word is pure Turkish.

News text summary study has some advantages when compared to other text summary studies. The news text is checked many times before it is published, and this prevents any typos [5]. In this way, a data set containing the news text gives better results for summarizing. The title is created according to the content of the news text, mostly reflecting the subject. It is important in the determination of word frequency or named entities. Keywords are used to access the news more easily or associate the news with another content. It also gives important clues about the news text such as the title. Named entities such as the names of institutions, organizations, and dates in the news text are frequently used in the content. This increases the frequency of the related word and may yield wrong results in some statistical methods. Generally, the number of sentences in the breaking news is fewer. The presence of fewer sentences than the minimum number of sentences required for the summary in the text prevents the formation of a good summary.

News content should be written in a way that every reader can easily understand. An easily understandable text for Turkish depends on pure Turkish words. The two-vowel harmony rule and hybrid model used in the detection of these words gave successful results for extractive summarization.

In this study, we present the structure of the sentence in Turkish news texts, and we use various statistical features for the extractive summary. We propose major vowel harmony and minor vowel harmony features and a hybrid model that has no literature studies.

This paper is organized as follows. In Section 2, we analyze previous works about extractive summarization. In Section 3, we explain studies on sentence ranking methods in the literature. In Section 4, we first give information about the dataset. Then we explain the preprocessing work on the dataset. We explain ROUGE and BLEU metrics used for summary evaluation and list the sentence ranking methods. In this section, we explain the two proposed methods and the hybrid method. In Section 4, we present the results obtained from the experimental study as a table. In Section 5, we list the general evaluation and provide future research directions.

Related Work

Automatic text summarization is a branch of Natural Language Processing. Many methods have been proposed in the literature to solve this problem. In this section, some of these studies are reviewed, and the studies are summarized in Table 1.

Yeh et al. proposed two new techniques for summarizing a text. Modified Corpus Based Approach (MCBA) and Text Relationship Map (TRM) technique based on Latent Semantic Analysis (LSA + TRM) [6]. MCBA is an instructible summarization approach and works on scoring sentences. The summarization process was carried out on features such as the position of the sentence, keywords, similarity to the title. In the study, the importance of sentence position and the Genetic Algorithm-based scoring process was performed. 100 political articles were used as a dataset. In the evaluation f-measure value were used, a result of 0.5151 was obtained.

Fattah and Ren suggested an approach that makes use of statistical features for summarization [7]. In this approach, statistical features such as similarity of words between paragraphs, the similarity of words between sentences, frequency of terms, location of the sentence, title are used. It uses a language-independent approach. The text features used are language independent. The features selected in the training phase are used in training the Naive-Bayes Classifier and Support Vector Machine. The weights of the features obtained from the training phase help to rank the importance of the sentences. The DUC 2002 dataset and the ROUGE-1 metric were used, and the result was 0.3862.

Ouyang et al. used Regression models to rank sentences in query-based summarization [8]. In this approach, features such as named entity matching, word matching, and semantic matching are used. Human summaries were used as training data. Afterward, various approaches based on the N-gram technique, which calculates the relevance scores of the sentences over these training data, were developed and compared. A mapping function is performed by collecting features of predefined sentences with the help of training data. The fitness of the sentences in the test data was estimated through the learned function. In the study on DUC 2005, DUC 2006, and DUC 2007 datasets, ROUGE-2 and ROUGE-SU4 metrics were used for evaluation. ROUGE-2 = 0.0757, ROUGE-SU4 = 0.1335 for DUC 2005 dataset; ROUGE-2 = 0.0926, ROUGE-SU4 = 0.1485 for DUC 2006 dataset; For the DUC 2007 dataset, ROUGE-2 = 0.1133, ROUGE-SU4 = 0.1652 values were obtained.

Baralis et al. proposed a new graph-based GRAPHSUM for summarization [9]. In this approach, data mining is used to discover correlations between multiple terms. Item sets with high correlations between terms were extracted from the data set and a correlation graph was created from these terms to determine the important sentences for the summary. The relationship at the graph nodes was estimated with a variant of the PageRank algorithm [10]. The sentences that best covered the correlation graph were selected for summary generation. Greedy algorithm was used for sentence selection. DUC 2004 dataset and ROUGE-2, ROUGE-SU4 metrics were used. Recall = 0.093, Precision = 0.099, F-measure = 0.097 on ROUGE-2; Recall = 0.015, Precision = 0.021, F-measure = 0.019 values were obtained on the ROUGE-SU4.

Alguliev et al. proposed an unsupervised summarization model that directly identifies important sentences from the document [11]. This approach is called Maximum Coverage and Minimum Redundancy (MCMR). In this method, three important features of a summary are optimized: relevance, redundancy, and length. A subset of the relevant text is selected from the document. The similarity was calculated using NGD-based similarity (Normalized Google Distance) and cosine similarity between the summarized text and the selected subset, and it was aimed to maximize this similarity. DUC 2005 and DUC 2007 datasets and ROUGE-2 and ROUGE-SU4 evaluation metrics were used. As a result, ROUGE-2 = 0.0790, ROUGE-SU4 = 0.1392 for DUC 2005, and ROUGE-2 = 0.1221, ROUGE-SU4 = 0.1753 for DUC 2007.

Ferreira et al. proposed a new graph-based clustering algorithm for summarization [12]. In the proposed algorithm, statistical and semantic similarities were studied. A four-dimensional graphical model was used. The vertices of the graph represent sentences, the edges of the graph represent the concepts of semantic similarity, statistical similarity, discourse relationships, and common reference resolution. The TextRank [13], [14] score was calculated for each node. Vertex with the highest TextRank score was chosen. The peaks were determined over the determined threshold value and each peak was configured to represent the cluster. In the selection of the sentences, the position information closest to the peak point was used. The DUC 2002 dataset and the F-measure metric were used. A value of 30% was obtained for 200 words and 25.4% for 400 words.

Kikuchi et al. proposed an approach for summarizing text using dependency between sentences and dependency between words [15]. Both dependencies are represented by two types of tree structures. A nested tree is created by replacing nodes in a document tree with a sentence tree. A subtree, whose nodes are random subtrees of the sentence tree, is extracted using this method. For summarization, the most important sentences were obtained by pruning the nested tree without losing the important content in the document. RST Discourse Treebank dataset and ROUGE-1 metric were used and the result was 0.354.

Liu et al. proposed MDS-Sparse, which performs summarization using the features of coverage, sparsity, and diversity [16]. At Level-1, the set of sentences in the summary text was sparsely represented by the original document set. At Level-2, sentences in the document are sparsely reconstructed by the summary set. Simulated annealing algorithm is used for this model. Each sentence in the document set is represented as a non-negative linear combination of only some summary sentences. This approach also aims to improve the linguistic quality of the abstract through rewriting. This method is a unified optimization framework based on compression. DUC 2006 and DUC 2007 datasets and ROUGE-1, ROUGE-2, ROUGE-SU4 metrics were used. For DUC 2006, ROUGE-1 = 0.34439, ROUGE-2 = 0.05122 and ROUGE-SU4 = 0.10717; For DUC 2007, the results ROUGE-1 = 0.35399, ROUGE-2 = 0.06448 and ROUGE-SU4 = 0.11669 were obtained.

Fang et al. suggested a summarization approach based on the topic [17]. Based on subject factors, various feature groups were extracted and used for sentence selection. A topic-based summary explains the different aspects. The proposed method is applied for text and image summarization. After extracting various types of features from the documents, feature groups were created. For text summarization, word frequency, the position of the sentence, and length of the sentence were used. A feature vector is made by combining the extracted features. Groups such as adjectives, adverbs, verbs, nouns, pronouns, prepositions, wh markers, symbols, numbers were created for the word feature. Greedy algorithm was used to generate the summary. DUC 2003 and DUC 2004 datasets and ROUGE-1 and ROUGE-L metrics were used. ROUGE-1 = 0.31990, ROUGE-L = 0.29389 for DUC 2003, ROUGE-1 = 0.33743, ROUGE-L = 0.30706 for DUC 2004 were obtained. Du and H. Huo suggested a new automatic summarzation model for news text [5]. It is based on fuzzy logic rules, multi-feature and genetic algorithm. In this method, important words such as time, place, character are selected first, and then each sentence is weighted with a genetic algorithm. For DUC 2002 dataset ROUGE-1 = 0,48 ROUGE-2 = 0,21 were obtained. Ahmad et al implemented three techniques for generating the extractive summary [14]. For CNN dataset, Recall = 0.61, Precision = 0.22, F-measure = 0.31, for BBC, Recall = 0.46, Precision = 0.54, F-measure = 0.49 were obtained. Jia et al proposed NLSSum for extractive summary [18]. For WikiLingua dataset F-measure = 0.31were obtained.

Table 1. Chronological comparison of recent extractive summarization approaches

Reference	Dataset	Evaluation Metrics	Result
Yeh et al. (2005) [6]	100 political articles	Precision Recall F-measure	Recall = Precision = F- measure = 0.5151
Fattah and Ren (2009) [7]	DUC 2002	ROUGE-1	R-1 = 0.3862
Ouyang et al. 2011) [8]	DUC 2005, DUC 2006 DUC 2007	ROUGE-2 ROUGE-SU4	For DUC 2005 R-2 = 0.0757, R-SU4 = 0.1335 For DUC 2006 R-2 = 0.0926, R-SU4 = 0.1485 For DUC 2007, R-2 = 0.1133, R-SU4 = 0.1652
Baralis et al. (2012) [9]	DUC 2004	ROUGE-2 ROUGE-SU4	For R-2, Recall = 0.093, Precision = 0.099, F-measure = 0.097 For R-SU4, Recall = 0.015, Precision = 0.021, F-measure = 0.019
Alguliev et al. (2013) [11]	DUC 2005 DUC 2007	ROUGE-2 ROUGE-SU4	For DUC 2005, R-2 = 0.0790, R-SU4 = 0.1392 For DUC 2007, R-2 = 0.1221,R-SU4 = 0.1753
Ferreira et al. (2013) [12]	DUC 2002	F-measure	For 200 word F-measure = 30% For 400 word F-measure = 25.4%
Kikuchi et al. (2014) [15]	RST Discourse Treebank	ROUGE-1	R-1 = 0.354
Liu et al. (2015) [16]	DUC 2006 DUC 2007	ROUGE-1 ROUGE-2 ROUGE-SU4	DUC 2006, R-1 = 0.34439, R-2 = 0.05122 and R-SU4 = 0.10717. DUC 2007, R-1 = 0.35399, R-2 = 0.06448 and R-SU4 = 0.11669
Fang et al. (2017) [17]	DUC 2003 DUC 2004	ROUGE-1 ROUGE-L	For DUC 2003 R-1 = 0.31990, R-L = 0.29389. For DUC 2004 R-1 = 0.33743, R-L = 0.30706
Du and H. Huo (2020) [5]	DUC 2002	ROUGE-1 ROUGE-2	ROUGE-1 = 0,48 ROUGE-2 = 0,21
Ahmad et al. (2021) [14]	CNN / BBC News Dataset	ROUGE-1	For CNN, Recall = 0.61, Precision = 0.22, F-measure = 0.31 For BBC, Recall = 0.46, Precision = 0.54, F-measure = 0.49
Jia et al. (2022) [18]	WikiLingua	ROUGE-L	F-measure = 0,31

MATERIALS AND METHODS

Dataset

We used a dataset prepared and made publicly available by Fırat University Big Data and Artificial Intelligence Laboratory [19]. It consists of 132,641 records, and each record contains news titles, news content, and several keywords. We selected 93,894 records from the dataset because some news content had fewer than 5 sentences. The reason is that in our experimental study, we selected summary rates of 20%, 30%, 40%, 50% and 60%. If a news text contains fewer than 5 sentences, the resultant summary would have fewer than 1 sentence, Therefore the one sentence required for the summary is considered to be the threshold value. As a result, we eliminated 38,747 news from the original dataset. Some characteristic features of the dataset are presented in Table 2.

Data Preprocessing

We used the Python programming language and the NLP library Natural Language Processing Toolkit (NLTK) [20] in Python, Python-based TRNLP [21] library for data preprocessing operation.

Firstly, we divided the whole text into sentences using NLTK and stored them in an array with their respective sentence positions because we used sentence position information in the last operation to preserve the flow of meaning.

After sentence segmentation, we used tokenization for dividing the text into words. We performed the tokenization process using NLTK when our experimental studies mainly progressed according to the words in the text.

We removed the stop words from the news content [22]. These are the words that do not give any additional meaning to the sentence, mostly used as conjunctions and prepositions [23]. Additionally, we discarded punctuation marks and words containing exclamation statements.

The words in the sentences have various meanings and structural differences according to the suffixes they take. It is advantageous both structurally and semantically to determine the stem of the words in the sentence without loss of meaning. Turkish is an agglutinative language, and the addition of different suffixes to the stem of the word causes

Table 2. Characteristic features of the dataset

Characteristic	Value
Number of documents	93.894
Minimum document length	5 sentences
Maximum document length	586 sentences
Average document length	14.76
Minimum document (5 sentences)	İstanbul'da dün öğlen saatlerinde başlayan sağanak yağmur etkisini sürdürüyor. Gece saatlerinde etkisi artıran sağanak yağmur sabah saatlerinde de devam etti. Yağmur nedeniyle Zeytinburnu sahilinden Marmara Denizine çamur aktı. İlçedeki dereden denize karışan çamur tabakasının yaklaşık 2 kilometre açıkta bekleyen demirli gemilerin olduğu bölüme kadar ilerlediği görüldü. Çamur tabakasının dere üzerinden denize karışması havadan drone ile de görüntülendi.
Maximum document (586 sentences) (Some of the sentences are presented)	Yıldırım ve İmamoğlunun İstanbul üzerine tartıştığı programda adaylaraeşit sürede sorular yöneltildi. Yıldırım ve İmamoğlu soruları cevaplamadan önce birbirlerine Babalar Günü hediyesi takdim etti. İsmail Küçükkaya ortak yayından sonra İmamoğlu ve Yıldırım aileleri ile beraber çektirdiği hatıra fotoğrafını paylaştı Tüm Türkiyenin izlediği yayından öne çıkan başlıklar şöyle; Soru Biz bu seçime niye giriyoruz. Binali Yıldırım Seçmenlerin suçu yok. oylar sayılırken garip işler oldu. YSK bu durmu değerlendirerek yenilenmesine karar verdi. Seçime gitmek istemedik. CHP bize bu konuda yardımcı olmadı. Keşke oylar tamamen sayılabilseydi. Vatandaşı yormuş olmayacaktık....

new meanings independent of the stem. This property can be considered a disadvantage for the Turkish language in the Natural Language Processing process. We used the Python-based TRNLP [15] library for word stem detection.

The preprocessing process is completed after the above-mentioned steps and is ready for the sentence score calculation operation. Thus, we did not need to do the preprocessing step during the experimental study of each feature.

Evaluation Metrics

In our experimental study, we used two very common metrics to measure how well each proposed method performs: ROUGE and BLEU scores.

ROUGE is a library developed to provide automatic evaluation between summaries, used to measure how well the summarized data fits the original text [24]. It contains different methods. The most widely used is the Rouge-N method. In this method, the suitability of a given summary to the desired summary value is determined by looking at the overlap rate of n-gram words. The resulting f-score is between 0 and 1, with the best summary close to 1 f-score.

BLEU score, proposed by IBM, measures the similarity of the system output with the "n" reference translations previously created by the translators [25]. The similarity is measured by matching words and phrases. It is based on precision calculation. The precision calculation is obtained by dividing the total number of words (unigrams) in the candidate sentence and in the reference sentence(s) by the total number of words in the candidate sentence. BLEU score is a numeric value between 0 and 1, with the best summary close to 1.

Sentence Ranking Features

1. Term frequency feature

The term frequency of a word or document frequency is generally used for feature extraction in the text [14], [26].

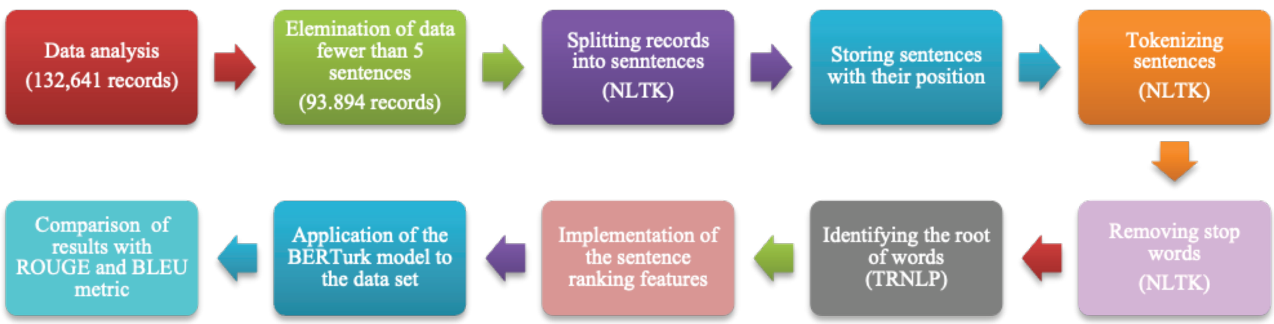

Figure 1. Overall process of data processing, sentence ranking, and evaluation.

The high frequency of the term indicates that the word and the sentences in which those words are used are important. Term weighting was originally conceived by Luhn [27]. Luhn proposed that a word can be weighted by its relative frequency in a given document. We calculated the term frequency in the sentences, then we ranked the sentences regarding the frequent occurrences of these terms. We added the sentences in the summary section according to the obtained sentence ranking and summary rate.

2. Keyword feature

In the dataset, there are a few keywords that qualify the news. We used these keywords like high-frequency words. We assumed that the keywords in the dataset were important words. We included these words in the high-frequency word group like term frequency feature and rated the sentences with this method. Finally, we added the sentences in the summary section according to the obtained sentence ranking and summary rate. This algorithm is similar to the topic model. Topic model programs assume that any document is composed by selecting words from possible topics. Topic model programs are created by statistically obtaining related words according to the topic in a document [28].

3. Title feature

In the dataset, every news has a title that contains one or more words. These words are significant for the content. The title is important because it reflects the scope of the news subject [29]. In this method, we used the frequency of the words in the title. We ranked the sentences with this method. Finally, we added the sentences in the summary section according to the obtained sentence ranking and summary rate.

4. Sentence position (first sentence) feature

Generally, the first sentence of the document gives important clues about the scope of the subject. The concept, place, date, person, and time information can be explained in this sentence [30]. In this feature we used this sentence and analyzed the document according to the words of the first sentence, then used the frequency of the words in the first sentence. We ranked the sentences regarding the frequent occurrences of these terms. We added the sentences in the summary section according to the obtained sentence ranking and summary rate.

5. Sentence position (last sentence) feature

Sometimes the last sentence in a document contains lots of information about the context. The importance of this sentence makes the words important. We used the words in the last sentence and considered these words to be high-frequency terms. Then we ranked other sentences regarding the frequent occurrences of last sentence words and finally, we initialized the summary according to the obtained sentence rating and summary rate.

6. Sentence length feature

We used this feature based on the number of words in the sentence. Using too many words in a sentence may result in using unnecessary or repetitive words. Similarly, having too few words in a sentence indicates that there is not enough information. We used a threshold value to select the sentence for summarization. We calculated the threshold value by averaging the maximum number of words in a sentence and the minimum number of words in a sentence. Finally, according to the number of sentences that should be in line with the summary rate, we included the sentences in the summary section, half of which is greater than the threshold value and the other half is less than the threshold value.

7. Named entity feature

In this method, some information gives important details like a proper name, institution, organization, date [31]. These sentences should be selected for the summarization because of their importance. We used the Python-based TRNLP [15] library to get named entities in sentences. We searched the document and compared a named entity dataset in Turkish and ranked every sentence according to the named entity count. We generated the summary according to sentence ranking.

8. Positive-negative feature

We used this method to detect the positive or negative property of words. We used the Python-based TRNLP [15] library to set a value between 0 and 1 to control expressions and affixes. This is a rule-based method, such as the word in Turkish "şekersiz" (sugar-free) being a negative expression in terms of not containing the relevant entity, and being a negative expression in terms of containing the verb in the word in Turkish "gelme" (don't come). We ranked the sentences with minimum negative expressions for the summarization. We added a summary according to sentence ranking and summarization rate. This method is somewhat similar to the aspect-base-based approach. It uses positive, negative and neutral approaches of words instead of sentences [32]. This approach can also be considered as a classification algorithm [22].

9. Major vowel harmony feature

Major vowel harmony is an important rule in the Turkish language and it distinguishes Turkish from many languages. In this rule, the major vowel harmony rule is explored in the placement of the eight vowel letters in the alphabet in the next syllables of the word. In Turkish, there are four back vowel letters (a, ı, o, u) and four front vowel letters (e, i, ö, ü). In the major vowel harmony rule, all these vowels are used. In this rule, a syllable that starts with a back vowel is followed by a back vowel in the same way. Likewise, the syllable that starts with the front vowel continues with the front vowel. The situation of vowel (back or front) in the first syllable of the words continues in the same way in the next syllables [33]. We used this feature for the root forms of the words in the news content and ranked the

Table 3. Major vowel harmony rule sample words

Word	Rule Control	Vowels
Ağaç (Tree)	Compliance	(a - a) vowels
Anahtar (Key)	Compliance	(a – a - a) vowels
Karanlık (Dark)	Compliance	(a – a - ı) vowels
Tiyatro (Theatre)	Non-compliance	(i – a - o) vowels
Trafik (Traffic)	Non-compliance	(a - i) vowels
İstasyon (Station)	Non-compliance	(i – a - o) vowels

Table 4. Minor vowel harmony rule sample words

Word	Rule Control	Vowels
Sorun (Problem)	Compliance	(o - u) vowels
Yuvarlak (Round)	Compliance	(u – a) vowels
Yarışma (Competition)	Compliance	(a –ı) vowels
Kavun (Melon)	Non-compliance	(a - u) vowels
Çamur (Mud)	Non-compliance	(a - u) vowels
Kavuşmak (Meet)	Non-compliance	(a – u - a) vowels

sentences according to the number of words matching this rule. Finally, according to the summarization rate and sentence ranking, we added sentences to the summary section. We presented some words that compliance and non- compliance with this rule as examples in Table 3.

10. Minor vowel harmony feature

This rule is used to check whether a word is pure Turkish. In this rule;

a) If there is a flat vowel (a, e, ı, i) in any syllable of the word, the next syllable must also contain flat vowels.
b) If there is a rounded vowel (o, ö, u, ü) in any syllable of the word, the next first syllable must have a wide flat (a,e) or narrow rounded (u, ü) [12].

We ranked the sentences according to the number of words matching this rule for the summarization. Finally, we initialized the summary section according to the summary rate and sentence ranking. We presented some words that compliance and non- compliance with this rule as examples in Table 4.

11. Major and minor vowel harmony feature

In this method, we used two features together. We ranked sentences according to the number of words matching major and minor vowel features together. We calculated sentence ranking using this hybrid method. We added sentences to the summary section according to the summary rate and sentence ranking.

EXPERIMENTAL RESULTS AND DISCUSSION

In this section, we provide experimental results of each feature in terms of ROUGE and BLEU scores. The results for term frequency, keyword feature, title feature, sentence location (first sentence), sentence location (last sentence), sentence length feature, named entity count feature, positive-negative feature, major vowel feature, minor vowel feature, and hybrid feature (major and minor vowel feature) are presented in Table 5, 6, 7, …. , 15, 16 respectively. We compared the three proposed methods with the BERTurk. Results are presented in Table 17. Additionally, we provide an in-depth analysis of and discussion of our findings in this section.

We used f-score, precision and recall values of ROUGE metric [34]. From the results given in their respective tables, we observe that as the summary rate increases, both Rouge-R, Rouge-F, and BLEU scores consistently increase. This is naturally expected because the number of sentences selected as part of the summary increases while the summary rate increases. On the other hand, Rouge-P scores are observed to decrease since as the summary rate increases, the universal cluster becomes larger and larger, which causes the number of overlapping n-grams to decrease proportionally. It has been observed that this part has decreased, albeit with a small value.

According to ROUGE and BLEU metrics, it was observed that the results of the summary between 20% and 40% progressed rapidly in a positive direction, the results after 40% increased when compared to the previous summary, but the same acceleration did not continue. This situation can be evaluated as the fact that the part of the text that can be a summary contains the same data after a certain period or that the sentences are rephrased in detail and do not bring additional information. The ROUGE-2 metric gave results close to the ROUGE-1 metric but lower value. This is because the ROUGE-2 metric checks for bigram-shaped phrases, while the words for ROUGE-1 are checked for overlapping independently and without any association (unigram). The ROUGE-L measures the longest common sequence (LCS) between the summary text and the original document. This gives the longest string of tokens shared between both. ROUGE-L value in all features gave better results than ROUGE-1 and ROUGE-2. This is because in the news text, important aspects of the event, named entities, place, time, etc. Because it contains information, it may be longer than other sentences. Finding these sentences by the observed features ensures that the longest common row is also of high value.

It has been observed that the accelerated increase in the term frequency feature is between 20% summary rate and 30% summary rate on all metrics. Although the amount to be included in the summary increased after this rate, there was no accelerated increase since the number of terms that could be considered important in the news text did not go beyond the scope of the event. Since the ROUGE-L metric uses the longest common sequence (LCS) value, it gave more successful results than ROUGE-1 and ROUGE-2. While BLEU results showed an accelerated increase between 20% summary rate and 30% summary rate, this acceleration was slower in the following rates. The results for the term frequency feature are presented in Table 5.

The keyword feature shows parallelism with the word frequency feature according to all metrics. Considering that this feature is a variant of the word frequency feature, it can be considered natural. Using the frequency of the keyword will not provide any advantages because the news texts are

Table 5. Term frequency feature ROUGE and BLEU metric results

Summary Rate	Rouge1-F	Rouge1-P	Rouge1-R	Rouge2-F	Rouge2-P	Rouge2-R	RougeL-F	RougeL-P	RougeL-R	BLEU
20	0.3960	0.9931	0.2539	0.3837	0.9797	0.2453	0.4413	0.9992	0.2904	0.06
30	0.5461	0.9793	0.3872	0.5303	0.9585	0.3749	0.5848	0.9990	0.4214	0.17
40	0.6251	0.9597	0.4727	0.6053	0.9351	0.4565	0.6608	0.9989	0.5013	0.26
50	0.6887	0.9263	0.5593	0.6644	0.8977	0.5382	0.7265	0.9987	0.5788	0.36
60	0.7380	0.8783	0.6467	0.7091	0.8462	0.6203	0.7860	0.9986	0.6552	0.46
AVERAGE	0.5988	0.9473	0.4639	0.5785	0.9234	0.4470	0.6399	0.9989	0.4894	0.26

Table 6. Keyword feature ROUGE and BLEU metric results

Summary Rate	Rouge1-F	Rouge1-P	Rouge1-R	Rouge2-F	Rouge2-P	Rouge2-R	RougeL-F	RougeL-P	RougeL-R	BLEU
20	0.3886	0.9969	0.2479	0.3765	0.9838	0.2394	0.4365	0.9991	0.2867	0.06
30	0.5378	0.9946	0.3766	0.5252	0.9792	0.3669	0.5812	0.9990	0.4179	0.16
40	0.6207	0.9938	0.4591	0.6085	0.9799	0.4491	0.6594	0.9989	0.4997	0.23
50	0.7022	0.9932	0.5513	0.6913	0.9817	0.5418	0.7348	0.9988	0.5888	0.33
60	0.7868	0.9926	0.6581	0.7779	0.9835	0.6498	0.8122	0.9987	0.6899	0.44
AVERAGE	0.6072	0.9942	0.4586	0.5959	0.9816	0.4494	0.6448	0.9989	0.4966	0.24

Table 7. Title feature ROUGE and BLEU metric results

Summary Rate	Rouge1-F	Rouge1-P	Rouge1-R	Rouge2-F	Rouge2-P	Rouge2-R	RougeL-F	RougeL-P	RougeL-R	BLEU
20	0.3965	0.9970	0.2541	0.3832	0.9811	0.2449	0.4415	0.9990	0.2910	0.06
30	0.5492	0.9951	0.3872	0.5342	0.9751	0.3757	0.5883	0.9989	0.4250	0.17
40	0.6335	0.9943	0.4726	0.6186	0.9760	0.4605	0.6680	0.9988	0.5094	0.25
50	0.7147	0.9936	0.5662	0.7011	0.9781	0.5545	0.7435	0.9987	0.5999	0.35
60	0.7978	0.9929	0.6728	0.7863	0.9805	0.6623	0.8201	0.9987	0.7009	0.46
AVERAGE	0.6184	0.9946	0.4706	0.6047	0.9782	0.4596	0.6523	0.9988	0.5052	0.26

written within certain patterns. Because the high-frequency word is usually defined as a keyword. The results for the keyword feature are presented in Table 6.

Based on the title feature, the term uses the frequency feature. Therefore, the results obtained are similar to term frequency and keyword feature. There are customized words in the title to reflect the document. These words often include named entities. The results for the title feature are presented in Table 7.

A balanced increase was observed for each summary rate in the sentence position (first sentence) feature. Generally, there is limited information to reflect the document in the first sentence and although the summary rate has increased, this restriction has increased to a certain extent, which is also reflected in the results. The results for the sentence location (first sentence) feature are presented in Table 8.

In the sentence position (last sentence) feature, low results were obtained for the 20% summary rate. For the number of 5 sentences determined as the threshold value, 1 sentence was not sufficient in terms of reflecting the information of the document. At this point, the fact that no additional information is usually given in the last sentences of the news texts supports this conclusion. Especially after the 30% summary rate, a balanced increase was observed. The sentence location feature (last sentence) did not produce as successful results as the sentence location feature (first sentence), which was caused by the difference between the first sentence and the last sentence in the news texts in terms of the information content and the news flow. The results for the sentence location (last sentence) feature are presented in Table 9.

When the results of the sentence length feature were examined, a balanced increase was observed as the summary rate increased. The results obtained for the BLEU metric were not as high as the ROUGE. Because the BLEU metric follows a more stringent method for n-gram measurements than the ROUGE metric. In the sentence length feature, the threshold value determined for each news text increases in parallel with the summary rate. With the increase in the summary rate, more sentences close to the threshold value were included in the summary, which resulted in a positive increase in metrics. The results for the sentence length feature are presented in Table 10.

A balanced increase in results was observed as the summary rate increased for the named entity count feature. Since news texts contain dense named entities, as the summary rate increases, more named entities will be observed that the metrics can compare, which will have a positive impact on the results. ROUGE-2 values were lower than ROUGE-1 values as in other features. The results for the named entity count feature are presented in Table 11.

In the positive-negative feature if a word is positive, it takes the value 1, if it is negative, it takes the value 0 and the words in the document are evaluated in this range. Since the news texts contain more positive words to reflect this feature, it has been observed that the results increase as the summary rate increases. As the summary rate increases, the number of words in the summary increases, and thus the

Table 8. Sentence location (first sentence) feature ROUGE and BLEU metric results

Summary Rate	Rouge1-F	Rouge1-P	Rouge1-R	Rouge2-F	Rouge2-P	Rouge2-R	RougeL-F	RougeL-P	RougeL-R	BLEU
20	0.3714	0.9469	0.2381	0.3509	0.9136	0.2242	0.4159	0.9993	0.2706	0.05
30	0.4917	0.8845	0.3494	0.4605	0.8345	0.3264	0.5337	0.9992	0.3746	0.15
40	0.5695	0.8804	0.4301	0.5372	0.8343	0.4050	0.6131	0.9992	0.4534	0.23
50	0.6429	0.8766	0.5177	0.6116	0.8362	0.4918	0.6896	0.9991	0.5377	0.32
60	0.7142	0.8697	0.6140	0.6851	0.8356	0.5885	0.7668	0.9991	0.6298	0.42
AVERAGE	0.5579	0.8916	0.4299	0.5291	0.8508	0.4072	0.6038	0.9992	0.4532	0.23

Table 9. Sentence location (last sentence) feature ROUGE and BLEU metric results

Smmary Rate	Rouge1-F	Rouge1-P	Rouge1-R	Rouge2-F	Rouge2-P	Rouge2-R	RougeL-F	RougeL-P	RougeL-R	BLEU
20	0.2637	0.8354	0.1644	0.2378	0.7750	0.1480	0.2938	0.9929	0.1804	0.03
30	0.5227	0.9961	0.3635	0.5035	0.9670	0.3495	0.5611	0.9991	0.3999	0.14
40	0.5692	0.9957	0.4095	0.5499	0.9680	0.3949	0.6057	0.9991	0.4458	0.23
50	0.6145	0.9952	0.4584	0.5955	0.9693	0.4436	0.6485	0.9991	0.4938	0.34
60	0.6600	0.9948	0.5116	0.6416	0.9707	0.4969	0.6911	0.9991	0.5453	0.46
AVERAGE	0.5260	0.9634	0.3815	0.5057	0.9300	0.3666	0.5600	0.9979	0.4130	0.24

measured word values are higher. The results for the positive-negative feature are presented in Table 12.

An increase in the summary rate ensures that there are more words in the abstract. In this way, there are more words in which major vowel harmony can be checked. Since the news texts are written in pure Turkish at a level that every reader can understand, more words that comply with this rule are observed. A balanced increase was observed in both ROUGE and BLEU metrics due to the increase in the summary ratio. ROUGE-2 value was close to ROUGE-1 but gave a low result. ROUGE-L value provided higher success than ROUGE-1 and ROUGE-2. The results for the major vowel harmony feature are presented in Table 13.

Table 10. Sentence length feature ROUGE and BLEU metric results

Summary Rate	Rouge1-F	Rouge1-P	Rouge1-R	Rouge2-F	Rouge2-P	Rouge2-R	RougeL-F	RougeL-P	RougeL-R	BLEU
20	0.3478	0.9990	0.2124	0.2913	0.9667	0.1729	0.3478	0.9989	0.2124	0.01
30	0.4817	0.9987	0.3203	0.4156	0.9525	0.2682	0.4816	0.9987	0.3203	0.05
40	0.5249	0.9987	0.3603	0.4573	0.9508	0.3046	0.5248	0.9986	0.3603	0.11
50	0.6037	0.9965	0.4378	0.5692	0.9452	0.4119	0.6513	0.9984	0.4887	0.31
60	0.7487	0.9985	0.6030	0.6809	0.945	0.536	0.7486	0.9984	0.6029	0.41
AVERAGE	0.5413	0.9982	0.3867	0.4828	0.9520	0.3387	0.5508	0.9986	0.3969	0.17

Table 11. Named entity count feature ROUGE and BLEU metric results

Summary Rate	Rouge1-F	Rouge1-P	Rouge1-R	Rouge2-F	Rouge2-P	Rouge2-R	RougeL-F	RougeL-P	RougeL-R	BLEU
20	0.4460	0.9964	0.2944	0.4328	0.9810	0.2848	0.4928	0.9991	0.3349	0.09
30	0.5925	0.9943	0.4295	0.5765	0.9733	0.4170	0.6327	0.9989	0.4704	0.21
40	0.6709	0.9937	0.5134	0.6544	0.9735	0.4998	0.7054	0.9989	0.5520	0.30
50	0.7448	0.9931	0.6029	0.7290	0.9750	0.5893	0.7731	0.9988	0.6371	0.39
60	0.8189	0.9924	0.7019	0.8052	0.9775	0.6894	0.8404	0.9987	0.7295	0.50
AVERAGE	0.6546	0.9940	0.5084	0.6396	0.9761	0.4961	0.6889	0.9989	0.5448	0.30

Table 12. Positive-negative feature ROUGE and BLEU metric results

Summary Rate	Rouge1-F	Rouge1-P	Rouge1-R	Rouge2-F	Rouge2-P	Rouge2-R	RougeL-F	RougeL-P	RougeL-R	BLEU
20	0.3673	0.9963	0.2311	0.3554	0.9828	0.2230	0.4191	0.9992	0.2722	0.04
30	0.5132	0.9939	0.3538	0.5028	0.9822	0.3457	0.5636	0.9991	0.4003	0.13
40	0.5964	0.9933	0.4335	0.5871	0.9840	0.4258	0.6423	0.9989	0.4805	0.20
50	0.6792	0.9928	0.5243	0.6717	0.9860	0.5175	0.7186	0.9988	0.5686	0.30
60	0.7682	0.9924	0.6335	0.7623	0.9870	0.6277	0.7993	0.9987	0.6720	0.41
AVERAGE	0.5849	0.9938	0.4352	0.5759	0.9844	0.4279	0.6286	0.9990	0.4787	0.22

Table 13. Major vowel harmony feature ROUGE and BLEU metric results

Smmary Rate	Rouge1-F	Rouge1-P	Rouge1-R	Rouge2-F	Rouge2-P	Rouge2-R	RougeL-F	RougeL-P	RougeL-R	BLEU
20	0.5162	0.9971	0.3539	0.5041	0.9845	0.3446	0.5644	0.9991	0.3997	0.14
30	0.6636	0.9955	0.5028	0.6485	0.9773	0.4904	0.7033	0.9991	0.5477	0.30
40	0.7392	0.9949	0.5928	0.7233	0.9765	0.5790	0.7716	0.9991	0.6331	0.40
50	0.8072	0.9942	0.6841	0.7914	0.9768	0.6699	0.8323	0.9991	0.7175	0.50
60	0.8720	0.9935	0.7796	0.8574	0.9780	0.7659	0.8899	0.9990	0.8044	0.59
AVERAGE	0.7197	0.9950	0.5827	0.7049	0.9786	0.5700	0.7523	0.9991	0.6205	0.39

The minor vowel harmony feature gave very close results to the large vowel harmony rule described earlier. Such results would be expected since in Turkish, whether the word is pure Turkish or not is checked first with major vowel harmony and then with minor vowel harmony. Due to the structure of the news texts, the words reflecting this rule were observed more with the increase in the summary rate. The results for the minor vowel harmony feature are presented in Table 14.

The hybrid model gave a similar but lower result to the independent results of both models. This small difference occurred because the number of words reflecting both rules was less than the number of words in the two independently controlled rules. A similar distribution was observed for the other two features for both metrics. The results for the major and minor vowel harmony (hybrid) feature are presented in Table 15.

Table 14. Minor vowel feature ROUGE and BLEU metric results

Summary Rate	Rouge1-F	Rouge1-P	Rouge1-R	Rouge2-F	Rouge2-P	Rouge2-R	RougeL-F	RougeL-P	RougeL-R	BLEU
20	0.5164	0.9971	0.3541	0.5043	0.9846	0.3448	0.5657	0.9991	0.4008	0.14
30	0.6637	0.9954	0.5029	0.6487	0.9773	0.4906	0.7045	0.9991	0.5491	0.30
40	0.7392	0.9948	0.5928	0.7234	0.9766	0.5791	0.7728	0.9991	0.6346	0.40
50	0.8071	0.9942	0.6839	0.7913	0.9768	0.6697	0.8333	0.9991	0.7190	0.50
60	0.8718	0.9935	0.7793	0.8572	0.9779	0.7656	0.8909	0.9990	0.8059	0.59
AVERAGE	0.7197	0.9950	0.5826	0.7050	0.9787	0.5700	0.7534	0.9991	0.6219	0.37

Table 15. Major and minor vowel (hybrid) feature ROUGE and BLEU metric results

Summary Rate	Rouge1-F	Rouge1-P	Rouge1-R	Rouge2-F	Rouge2-P	Rouge2-R	RougeL-F	RougeL-P	RougeL-R	BLEU
20	0.5112	0.9972	0.3496	0.4990	0.9845	0.3403	0.5609	0.9991	0.3965	0.14
30	0.6588	0.9955	0.4977	0.6437	0.9773	0.4853	0.7001	0.9991	0.5441	0.29
40	0.7346	0.9949	0.5872	0.7187	0.9765	0.5735	0.7687	0.9991	0.6294	0.39
50	0.8031	0.9943	0.6785	0.7872	0.9768	0.6643	0.8298	0.9991	0.7141	0.49
60	0.8686	0.9936	0.7743	0.8539	0.9780	0.7606	0.8880	0.9990	0.8014	0.58
AVERAGE	0.7153	0.9951	0.5774	0.7005	0.9786	0.5648	0.7495	0.9991	0.6171	0.38

Table 16. All feature ROUGE and BLEU metric average results

Feature	Rouge1-F	Rouge1-P	Rouge1-R	Rouge2-F	Rouge2-P	Rouge2-R	RougeL-F	RougeL-P	RougeL-R	BLEU
Term Frequency Feature	0.5988	0.9473	0.4639	0.5785	0.9234	0.447	0.6399	0.9989	0.4894	0.26
Keyword Feature	0.6072	0.9942	0.4586	0.5959	0.9816	0.4494	0.6448	0.9989	0.4966	0.24
Title Feature	0.6184	0.9946	0.4706	0.6047	0.9782	0.4596	0.6523	0.9988	0.5052	0.26
First Sentence Feature	0.5579	0.8916	0.4299	0.5291	0.8508	0.4072	0.6038	0.9992	0.4532	0.23
Last Sentence Feature	0.5260	0.9634	0.3815	0.5057	0.9300	0.3666	0.5600	0.9979	0.413	0.24
Sentence Length Feature	0.5413	0.9982	0.3867	0.4828	0.952	0.3387	0.5508	0.9986	0.3969	0.17
Named Entity Feature	0.6546	0.9940	0.5084	0.6396	0.9761	0.4961	0.6889	0.9989	0.5448	0.30
Positive-Negative Feature	0.5849	0.9938	0.4352	0.5759	0.9844	0.4279	0.6286	0.999	0.4787	0.22
Major Vowel Feature	**0.7197**	**0.9950**	**0.5827**	**0.7049**	**0.9786**	**0.5700**	**0.7523**	**0.9991**	**0.6205**	**0.39**
Minor Vowel Feature	**0.7197**	**0.9950**	**0.5826**	**0.7050**	**0.9787**	**0.5700**	**0.7534**	**0.9991**	**0.6219**	**0.37**
Major And Minor Vowel Feature	**0.7153**	**0.9951**	**0.5774**	**0.7005**	**0.9786**	**0.5648**	**0.7495**	**0.9991**	**0.6171**	**0.38**

Table 17. Major, minor, hybrid model and BERTurk summarizer results

Feature	Rouge1-F	Rouge1-P	Rouge1-R	Rouge2-F	Rouge2-P	Rouge2-R	RougeL-F	RougeL-P	RougeL-R	BLEU
BertTURK	0.7636	0.9976	0.6992	0.8458	0.9829	0.684	0.9027	0.9997	0.7446	0.46
Major Vowel Feature	0.7197	0.9950	0.5827	0.7049	0.9786	0.5700	0.7523	0.9991	0.6205	0.39
Minor Vowel Feature	0.7197	0.9950	0.5826	0.7050	0.9787	0.5700	0.7534	0.9991	0.6219	0.37
Major And Minor Vowel Feature	0.7153	0.9951	0.5774	0.7005	0.9786	0.5648	0.7495	0.9991	0.6171	0.38

The mean values of all features are presented in Table 16 for comparison between features. Major vowel harmony feature, minor vowel harmony feature, and a hybrid model of these two features outperformed the other eight features. These results were observed in both the ROUGE and BLEU metrics.

The named entity feature had the closest result to the three proposed features. It is natural to have high performance for the summary, as elements such as place, time, institution name, which are named assets, are often used in news texts. In the three proposed methods, approximately 0.71 f-score was obtained in the ROUGE-1 metric, while the result was 0.65 in the named entity feature. For the three methods, approximately 0.70 f-score was obtained in the ROUGE-2 metric, while 0.63 f-score, approximately 0.75 f-score was obtained in the ROUGE-L metric, while 0.68 f-scores was obtained in the named entity feature. These results obtained for ROUGE show parallelism for BLEU as well. While values between 0.37 and 0.39 were observed for the three methods, a result of 0.3 was obtained for the named entity property that received the closest result.

We compared the three proposed methods with the BERTurk [35] model prepared for Turkish based on Google BERT [36]. Google BERT (Bidirectional Encoder Representations from Transformers) is a state-of-art method that sorts the most consistent and accurate results by taking all the words into a logical evaluation instead of processing the words separately in the text. The results obtained gave very close results to this state-of-the-art method and showed that it is worth developing. The results are presented in Table 17.

CONCLUSION

In this study, we conducted an experimental study on the features used in sentence grading and proposed new methods based on major vowel harmony and minor vowel harmony features for the extraction summarization in Turkish. In addition to the sentence grading features that are frequently used in the literature, we used the features of major vowel harmony, minor vowel harmony, and hybrid method of these methods. None of these methods were used before. As a result, we observed that these features can compete with other features. In our work with the Python programming language, we carried out all the features independently of each other after preprocessing operation. We evaluated summaries using ROUGE and BLEU metrics.

Both ROUGE and BLEU metrics yielded similar results for sentence ranking methods. In all ranking methods, we observed that as the summary rate increased, the metric results increased positively. Because the increase in the summary rate allows more sentences to be included in the summary section, and thus produces positive results in the comparison of the main text and the metrics. We compared the three proposed methods with the BERTurk. The results obtained gave very close results to this state-of-the-art method and showed that it is worth developing.

In this study, two rules in Turkish were tested on the roots of words. We think that the reason why these two methods and the hybrid method achieved high value in this paper is that this rule is not sought in non-Turkish words and applied over pure Turkish words. We observe that simple and pure Turkish words are intensively used in the news texts because such texts are written in a way that can easily be understood by the reader. The sentences which contain such words are included in the summary more often because they are the important sentences in the main text.

There were limitations in the study due to the agglutinative language structure of Turkish. The fact that there are too many words in the structure and at the end of the word has caused difficulties in determining the root of the word. This affects the semantic context. Working with a larger data set will yield better results.

We propose to consider the use of two features and hybrid features and feature weighting method over artificial neural network in future studies. A hybrid model can be created by using a structure such as Google BERT, which has gained popularity in recent years, and good work can be done for Turkish. In addition, on a larger-scale data set, assessments can be made based on gold summaries and supervised learning.

ACKNOWLEDGEMENTS

This study is derived from the Ph.D. thesis titled Extractive Based Automatic Text Summarization in Turkish Texts.

AUTHORSHIP CONTRIBUTIONS

Authors equally contributed to this work.

DATA AVAILABILITY STATEMENT

The authors confirm that the data that supports the findings of this study are available within the article. Raw data that support the finding of this study are available from the corresponding author, upon reasonable request.

CONFLICT OF INTEREST

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

ETHICS

There are no ethical issues with the publication of this manuscript.

REFERENCES

[1] Ong WH,. Tay KG, Chew CC, Huong A. A comparative study of extractive summary algorithms using natural language processing. In: 2020 IEEE Student Conference on Research and Development (SCOReD);2020 Sep; Batu Pahat, Malaysia. 2020. pp. 406–410. [CrossRef]

[2] Asa AS, Akter S, Uddin P, Hossain D, Roy SK, Afjal MI. a comprehensive survey on extractive text summarization techniques. Am J Eng Res 2017;6:226–239.

[3] Chen KY, Liu J, Chen B, Wang HM, Jan EE, Hsu WL, et al. Extractive broadcast news summarization leveraging recurrent neural network language modeling techniques. IEEE/ACM Trans Audio Speech Lang 2015;23:1322–1334. [CrossRef]

[4] Güran A, Bayazit NG, Gürbüz MZ. Efficient feature integration with Wikipedia-based semantic feature extraction for Turkish text summarization. 2013;21:1411–1425.

[5] Du Y, Huo H. news text summarization based on multi-feature and fuzzy logic. IEEE Access 2020;8:140261–140272. [CrossRef]

[6] Yeh JY, Ke HR, Yang WP, Meng IH. Text summarization using a trainable summarizer and latent semantic analysis Inform Process Manag 2005;41:75–95. [CrossRef]

[7] Fattah MA, Ren F. GA, MR, FFNN, PNN and GMM based models for automatic text summarization. Comput Speech Lang 2009;23:126–144. [CrossRef]

[8] Ouyang Y, Li W, Li S, Lu Q. Applying regression models to query-focused multi-document summarization. Inf Process Manag 2011;47:227–237. [CrossRef]

[9] Baralis E, Cagliero L, Jabeen S, Fiori A. Multi-document summarization exploiting frequent itemsets. In: Proceedings of the 27th Annual ACM Symposium on Applied Computing - SAC '12; 2012; Trento, Italy. 2012. p. 782. [CrossRef]

[10] Page L, Brin S, Motwani R, Winograd T. The pagerank citation ranking: Bringing order to the web. Available at: https://mail.google.com/mail/u/0/#inbox/FMfcgzGxSHgkqCDCgmQbPsctLcZlJFlf? projector=1&messagePartId=0.4 Accessed on Mar 13, 2024.

[11] Alguliev RM, Aliguliyev RM, Isazade NR. Multiple documents summarization based on evolutionary optimization algorithm. Expert Syst Appl 2013;40:1675–1689. [CrossRef]

[12] Ferreira R, Cabral L, Lins RD, Silva G, Cavalcanti GDC, Lima R, et al. Assessing sentence scoring techniques for extractive text summarization. Expert Syst Appl 2013;40:5755–5764. [CrossRef]

[13] Mihalcea R, Tarau P. TextRank: Bringing Order into Text. In Lin D, Wu D, editors. Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing; 2004 Jul; Bercelona, Spain: Association for Computational Linguistics;2004. pp. 404–411.

[14] Ahmad T, Ahmed S, Ahmad N, Aziz A, Mukul L. News article summarization: Analysis and experiments on basic extractive algorithms. Int J Grid Distrib Comput 2021;13:2366–2379.

[15] Kikuchi Y, Hirao T, Takamura H, Okumura M, Nagata M. Single Document Summarization based on Nested Tree Structure. In Proceedings of the 52nd Annual Meeting; 2014 Jun; Baltimore, Maryland: the Association for Computational Linguistics; 2014. p. 315–320. [CrossRef]

[16] Liu H, Yu H, Deng ZH. Multi-document summarization based on two-level sparse representation model. Knowl Inf Syst 2017;53:297–336. [CrossRef]

[17] Fang C, Mu D, Deng Z, Wu Z. Word-sentence co-ranking for automatic extractive text summarization. Expert Syst Appl 2016;72:189–195. [CrossRef]

[18] Jia R, Zhang X, Cao Y, Wang S, Lin Z, Wei F. Neural label search for zero-shot multi-lingual extractive summarization. Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Dublin, Ireland, 2022.

[19] Veri setleri - Büyük veri ve yapay zeka laboratuvarı. Available at: http://buyukveri.firat.edu.tr/veri-setleri/ Accessed on Mar13, 2024. [Turkish]

[20] Natural language toolkit - NLTK 3.5 documentation. Available at: https://www.nltk.org/ Accessed on Mar 13, 2024.

[21] Python Türkçe doğal dil işleme - Turkısh nlp, Türkçe için doğal dil işleme araçları. Available at: https://pypi.org/project/trnlp/ Accessed on Mar 14, 2024. [Turkish]
[22] Altan Z. A Turkish automatic text summarization system. In Proceedings of The Iasted International Conference on Artifical Intelligence and Applications; 2004 Feb 16-18; Innsbruck, Austira: 2004. p. 311-316.
[23] Hatipoğlu A, Omurca Sİ. A hybrid modelling for Turkish text summarization. Dokuz Eylül Univ Eng Fac J Sci Eng 2015;17:95–108.
[24] Lin CY, ROUGE: A package for automatic evaluation of summaries. In Proceedings of the Workshop on Text Summarization Branches Out; 2004 Jul; Barcelona, Spain: Association for Computational Linguistics; 2004. p. 74-81.
[25] Papineni K, Roukos S, Ward T, Zhu WJ. Bleu: A method for automatic evaluation of machine translation. In Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics; 2002 Jul 7-12; Philadelphia, Pennsylvania, USA. 2002. pp. 311-318. [CrossRef]
[26] Azam N, Yao J. Comparison of term frequency and document frequency based feature selection metrics in text categorization. Expert Syst Appl 2012;39:4760–4768. [CrossRef]
[27] Luhn HP. The automatic creation of literature abstracts. IBM J Res Dev 1958;2:159–165. [CrossRef]
[28] Akhtar N, Javed H, Ahmad T. Hierarchical summarization of text documents using topic modeling and formal concept analysis. In: Balas V, Sharma N, Chakrabarti A, editors. Data Management, Analytics and Innovation. Advances in Intelligent Systems and Computing; 2018; Singapore: Sprinter; 2018. pp. 21-33. [CrossRef]
[29] Turan M, Pek M. An extensible model for Turkish single document summarization. J CIEES 2022;2:16–20. [CrossRef]
[30] Kutlu M, Cigir C, Cicekli I. Generic text summarization for Turkish. J Comput 2010;58:1315–1323. [CrossRef]
[31] Kartal YS, Kutlu M. Machine learning based text summarization for Turkish news. In 2020 28th Signal Processing and Communications Applications Conference (SIU);2020 Oct; Gaziantep, Turkey: 2020. p. 1-4. [CrossRef]
[32] Akhtar N, Zubair N, Kumar A, Ahmad T. Aspect based sentiment oriented summarization of hotel reviews. Procedia Comput Sci 2017;115:563–571. [CrossRef]
[33] Tepeli Y, Arıcı AF. A study on vowel-consonant harmony violating words in Turkey Turkish. Int J Soc Educ Sci 2018;5:249–262. [CrossRef]
[34] Chakraborty R, Bhavsar M, Dandapat SK, Chandra J. Tweet summarization of news articles: An objective ordering-based perspective. IEEE Trans Comput Soc Syst 2019;6:761–777. [CrossRef]
[35] Schweter S. BERTurk-bert models for Turkish. Available at: https://zenodo.org/records/3770924 Accessed on Mar 13, 2024.
[36] Devlin J, Chang MW, Lee K, Toutanova K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In: Proceedings of NAACL-HLT 2019; 2019 Jun 2-7; Minneapolis, Minnesota, USA: 2019. p. 4171-4186.