ITC 1/46

Journal of Information Technology and Control
Vol. 46 / No. 1 / 2017
pp. 53-69
DOI 10.5755/j01.itc.46.1.13051

An Additive FAHP Based Sentence Score Function for Text Summarization

Received 2016/09/04 | Accepted after revision 2017/01/26

http://dx.doi.org/10.5755/j01.itc.46.1.13051

An Additive FAHP Based Sentence Score Function for Text Summarization

Aysun Güran, Mitat Uysal

Dogus University, Department of Computer Engineering, Acıbadem, Kadıköy, 34722 İstanbul
e-mails: adogrusoz@dogus.edu.tr, e-mail: muysal@dogus.edu.tr

Yeliz Ekinci

Istanbul Bilgi University, Department of Business Administration, Eyüp, 34060 İstanbul
e-mail: yeliz.ekinci@bilgi.edu.tr

Celal Barkan Güran

Istanbul Technical University, Department of Management Engineering, Maçka, 34357 İstanbul
e-mail: guranc@itu.edu.tr

Corresponding author: adogrusoz@dogus.edu.tr

This study proposes a novel additive Fuzzy Analytical Hierarchy Process (FAHP) based sentence score function for Automatic Text Summarization (ATS), which is a method to handle growing amounts of textual data. ATS aims to reduce the size of a text while covering the important points in the text. For this aim, this study uses some sentence features, combines these features by an additive score function using some specific weights and produces a sentence score function. The weights of the features are determined by FAHP – specifically Fuzzy Extend Analysis (FEA), which allows the human involvement in the process, uses pair-wise comparisons, addresses uncertainty and allows a hierarchy composed of main features and sub-features. The sentences are ranked according to their score function values and the highest scored sentences are extracted to create summary documents. Performance evaluation is based on the sentence coverage among the summaries generated by human and the proposed method. In order to see the performance of the proposed system, two different Turkish datasets are used and as a performance measure, the F-measure is used. The proposed method is compared with a heuristic algorithm, namely Genetic Algorithm (GA). Resulting performance improvements show that the proposed model will be useful for both researchers and practitioners working in this research area.

KEYWORDS: text summarization, fuzzy analytical hierarchy process, sentence score function.

Introduction

Automatic Text Summarization (ATS) is one of the most important ways to handle growing amounts of textual data. ATS sets the goal at reducing the size of a text while covering the important points in the text. ATS process consists of two types of summarization: abstractive and extractive. The abstractive summarization involves generating new sentences from given documents, whereas the extractive summarization attempts to identify the most important sentences for the overall understanding of a document. Most of the work in the literature is about extractive summarization due to its feasibility. These studies generally use some sentence features and combine these features with some specific weights to produce a sentence score function. The weights of the features can be acquired by either automatic (i.e. supervised learning methods and heuristic algorithms) or manual (i.e. human involved methods) techniques. Both of the techniques generate domain-dependent feature weights, which means that if dataset is changed, then the produced weights have to be recomputed. The automatic techniques are based on supervised learning and generally use heuristic algorithms such as genetic algorithms (GAs) [2, 27], particle swarm optimization (PSO) [3] and artificial bee colony (ABC) algorithm [15]. However, one disadvantage of the supervised methods is that they have to deal with training and testing phases of given dataset. On the other hand, manual techniques [15, 36, 37] decide the weights of features according to expert opinion. A significant advantage of the manual techniques is allowing the human involvement in the weight determination stage; incorporating expert knowledge and opinion in the problem. For both techniques, after deciding the proper weights for the evaluated sentence features, ATS ranks the sentences according to their score function's values and finally extracts the highest scored sentences to create summary documents.

In this study, we focus on an extractive summarization system and present a fuzzy analytic hierarchical process (FAHP) technique for weight calculation of the sentence features. The FAHP method is developed from the analytic hierarchical process (AHP). AHP is accepted as the best structural algorithm if the problem can be solved by pair-wise comparison and any criterion is not involved in interaction with another criterion [2]. In spite of the popularity of AHP, this method is often criticized for its inability to adequately handle the inherent uncertainty and imprecision associated with the mapping of the decision-maker's perception to exact numbers. Fuzzy set theory is used in this study to address the uncertainty. Fuzzy logic is capable of supporting human type reasoning in natural form. It has been seen as the most popular and easiest way to capture and represent fuzzy, vague, imprecise and uncertain domain knowledge in recent years [10]. These facts and definitions motivate us to incorporate fuzzy set theory with AHP method to solve the problem of determining the sentence feature weights. Since FAHP method is based on pair-wise comparison of the criteria under consideration, in this study the weight calculation is performed by the pairwise comparison of sentence features addressing the uncertainty in the expert judgments.

Many fuzzy methods and applications are presented by various authors. One of the best known of these methods is Fuzzy Extend Analysis (FEA) proposed by [8]. In our study the FEA is used to evaluate the sentence score function. In order to see the performance of the proposed system, Turkish datasets are used. Turkish datasets contain two different sets of documents. The first set involves 130 documents related to different areas and a human-generated extractive summary corpus. The second set contains 20 documents and 30 extractive summary corpora, which are prepared by 30 different assessors. The performance analysis of the proposed FAHP system is conducted on the human-generated summary corpora. As a performance measure, we use the F-measure score that determines the coverage between the manually and automatically generated summaries. In order to show the effectiveness of the proposed method, the results are compared with the results of a meta-heuristic; Genetic Algorithm (GA) based sentence combining method. As stated above, GA is used in this research area and usually gives competitive results with other meta-heuristic techniques such as PSO and ABC [4, 27]. Therefore, we choose GA for benchmark. Although meta-heuristic techniques do not adequately handle the inherent uncertainty and do not allow the human involvement in the weight determination stage, we wanted to make a benchmark since these techniques are widely used in the literature.

The remaining parts of the paper are organized as follows: Section 2 explains literature review; Section 3 outlines sentence features; Section 4 points out how sentence features are combined via the proposed FAHP based system. Section 5 explains combining the sentence features by GA. Section 6 presents the data corpus and the evaluation dataset. Section 7 presents the experimental results and finally Section 8 gives concluding remarks.

Literature review

Although researches on automatic text summarization started over 50 years ago, in the light of recently developed technology and improved natural language processing techniques, the field has been still very popular. In this section, a review of literature on extractive automatic ATS and FAHP will be presented.

Literature review for ATS

Creating extractive summary documents requires the selection of the most representative sentences of given documents. In literature there are lots of studies which analyse structural and semantic features of documents. These studies tend to represent information in terms of shallow features that are then selectively combined to yield a function used to extract representative sentences. These features include "term frequency", "sentence length", "location", "title feature", "cue words and phrases", "N-gram words", "some punctuation marks", "centrality of sentences", "similarity to other sentences", "name entities", "numerical data", etc. The study [25] created the first summarization system. This system points out that the frequency of word occurrence in a document provides a useful measure of word significance. The theoretical foundation for this model is provided by Zipf's Laws [43], which suggest that there is a power law relationship between the frequency of word occurrences and the rank of terms in a frequency table. Other studies that use different shallow features are: [11, 19, 30, 32, 33, 40].

The studies [5, 14, 16, 17 ,18, 24, 28, 35] represent documents with semantic sentence features based on Latent Semantic Analysis (LSA), Probabilistic Latent Semantic Analysis (PLSA) and Non-Negative Matrix Factorization (NMF), which analyze the relationships between a set of sentences and terms by producing a set of topics related to the sentences and the terms.

It is possible to combine sentence features with different techniques according to a hybrid system. The studies [6, 23] combine sentence features by using a fuzzy logic based hybrid system, whereas the studies [4, 38] use genetic algorithm based hybrid systems. Among the hybrid systems, GA is a widely used technique. GA is an evolutionary optimizer that takes a sample of possible solutions and employs mutation, crossover, and selection as the primary operators for optimization [13, 34]. Optimization-based methods have also been studied in the literature. The study [12] defines text summarization as a maximum coverage problem whereas the study [26] formalizes it as a knapsack problem. The study [3] models document summarization as a nonlinear 0-1 programming problem that covers main content of given documents through sentence assignment.

Literature review for FAHP

Multi Criteria Decision Making (MCDM) refers to find the best opinion from all of the feasible alternatives in the presence of multiple, usually conflicting, decision criteria. If the MCDM methodology is to be used in group decision-making, the analytic hierarchy process (AHP) is one of the best choices. Central to the resolution of a multi-criteria problem by the AHP is the process of determining the weights of the criteria and the final solution weights of the alternatives with respect to the criteria. As the true weights are unknown, they must be approximated. AHP elicits the decision maker's judgment of elements in a hierarchy and mathematically manipulates them to obtain the final preference weights of the decision alternatives with respect to the overall goal. On the other hand, AHP is often criticized for its inability to adequately handle the uncertainty and imprecision associated with the mapping of the decision-maker's perception to exact numbers. Moving from this point, fuzzy AHP (FAHP) is proposed by the researchers in order to incorporate uncertainty in the decision making problem. The fuzzy AHP technique can be viewed as an advanced analytical method developed from the traditional AHP. Despite the convenience of AHP in handling both quantitative and qualitative criteria of multi-criteria decision making problems based on decision makers' judgments, fuzziness and vagueness existing in many decision-making problems may contribute to the imprecise judgments of decision makers in conventional AHP approaches [20]. The reason for using fuzzy sets theory which was introduced by [41] is that it can deal with situations characterized by imprecision due to subjective and qualitative evaluations rather than to the effect of uncontrollable events on different variables. Imprecision is accommodated by possibility rather than probability distributions [31]. The study [22] listed three main reasons for incorporating fuzzy set theory in decision making: (i) imprecision and vagueness are inherent to the decision maker's mental model of the problem under study, (ii) the information required to formulate a model's parameters may be vague or not precisely measurable, (iii) imprecision and vagueness as a result of personal bias and subjective opinion may further dampen the quality and quantity of available information

In Fuzzy AHP (FAHP) method, the fuzzy comparison ratios are used to be able to tolerate vagueness. In the literature, there are several studies that use FAHP for decision making. For instance, the study [7] uses FAHP in order to select the universal provider considering the risk factors. The study [29] develops a new Fuzzy AHP based decision model which is proposed to select a Database Management System easily. The study [9] describes the design of a fuzzy decision support system in multi-criteria analysis approach for selecting the best plan alternative or strategy in environment watershed. In these studies AHP and FAHP are used to select the best alternative among many, using different criteria. Although AHP and FAHP calculate both the weights of the criteria and the alternatives, most of the studies use AHP for only weight calculation. The study [21] is one of the studies which use FAHP for weight calculation. However, to the best of our knowledge, there are very few studies that use AHP for text mining application. The study [15] is the first to use AHP in the area of Turkish text summarization. They combine sentence features with an AHP and artificial bee colony algorithm based hybrid system. Later on, [36] uses AHP techniques for Persian summarization. AHP based system doesn't require a training phase of a corpus. This can be regarded as an advantage since a training phase takes quite long time for algorithms to be executed. This fact motivates us to use an AHP-based method for text summarization beside the other advantages of manual techniques in the feature weight determination stage. In order to address the uncertainty in this method, we propose to use FAHP.

In this work we combine almost all sentence features that were previously used by many researches and we analyse the effects of FAHP method on text summarization task at the creation of overall sentence score function phase.

Sentence features

In this study, each sentence is represented as a feature vector formed of 15 features extracted from the document. We group the text features into five classes according to their level of text analysis. Table 1 shows the features and their classes.

These features are identified after the preprocessing of the original documents is done, like stemming and removing stop words. For stemming, Zemberek software [42] is used.

Table 1

Description of features

Goal	Classes	Features
Text summarization	F_1: Location knowledge	f_{11}: Sentence location f_{12}: Distributional features
	F_2: Similarity to main sentences	f_{21}: Similarity to first sentence f_{22}: Similarity to last sentence f_{23}: Similarity to title sentence
	F_3: Term frequency knowledge	f_{31}: Sentence length f_{32}: Term frequency f_{33}: Word sentence score f_{34}: Average Tf-Idf
	F_4: Thematic features	f_{41}: Numerical data f_{42}: Punctuation marks f_{43}: Positive key words f_{44}: Noun phrases
	F_5: Semantic features	f_{51}: LSA based features f_{52}: Centrality

f_{11}: Sentence location: Sentences at the beginning of documents always introduce the main topics that the documents discuss. To capture the significances of different sentence positions, each sentence in a document is given a rank according to (1):

$$Score_{f_{11}}(S_i) = \frac{N - P_i}{N} \quad (1)$$

where $\forall S_i \in d$ and P_i is the position of the i^{th} sentence and N is the total number of sentences of the document.

f_{12}: Distributional features: Term frequency can be regarded as a value that measures the importance of a term in a document. The importance of a term can be measured not only by its frequency but also by the compactness of its distribution. The study [39] proposes to use new features, connected with the distribution of terms within the document, called distributional features to categorize the text documents. They presented three different distributional features that point out the compactness of appearances of a term: $ComPact_{PartNum}$, $ComPact_{FLDist}$, $ComPact_{PosVar}$.

Let the distributional array of a term t_k is the $aray(t_k, d) = \{c_1, c_2, ..., c_N\}$, where c_N is the frequency of the term t_k in S_N. Then, values of $ComPact_{PartNum}$, $ComPact_{FLDist}$, and $ComPact_{PosVar}$ are defined as follows:

- $ComPact_{PartNum}$ denotes the number of parts where a term appears. This measure can be used to determine whether a term is compact or not since a term is less compact if it appears in different parts of a document. $ComPact_{PartNum}$ is computed as shown in (2):

$$ComPact_{PartNum}(t_k, d) = \sum_{i=1}^{N} c_i \succ 0 \ ?\ 1 : 0 \quad (2)$$

The expression c_i>0?1:0 is a conditional expression which means that if c_i is greater than zero then c_i=1 otherwise c_i=0.

- $ComPact_{FLDist}$ denotes the distance between a term's first and last appearance. This is also a pointer to the compactness of a term by the fact that, for a less compact term, the distance between the first mention and the last mention should be long. $ComPact_{FLDist}$ is computed as shown in (3):

$$\begin{aligned} &ComPact_{FLDist}(t_k, d) = Last_{App}(t_k, d) - First_{App}(t_k, d) \\ &First_{App}(t_k, d) = min_{i \in \{1..N\}} c_i \succ 0 ? i : N \\ &Last_{App}(t_k, d) = max_{i \in \{1..N\}} c_i \succ 0 ? i : -1 \end{aligned} \quad (3)$$

- $ComPact_{PosVar}$ denotes the variance of positions of all appearances. It is the mean of the product of position and the deviation of the position from the mean position (centroid) of all appearances. $ComPact_{PosVar}$ is computed as in (4):

$$\begin{aligned} &ComPact_{PosVar}(t_k, d) = \frac{\sum_{i=1}^{N} c_i * |i - centroid(t_k, d)|}{count(t_k, d)} \\ &count(t_k, d) = \sum_{i=1}^{N} c_i \\ &centroid(t_k, d) = \frac{\sum_{i=1}^{N} c_i * i}{count(t_k, d)} \end{aligned} \quad (4)$$

We base one of our features on this study and adapt the use of distributional features to text summarization. For the sentence S_i in the document *d*, the distributional feature score is computed according to (5):

$$Score_{f_{12}}(S_i) = \sum_{k=1}^{m_i} \begin{pmatrix} ComPact_{PartNum}(t_k, d) \\ + \\ ComPact_{FLDist}(t_k, d) \\ + \\ ComPact_{PasVar}(t_k, d) \end{pmatrix} \quad (5)$$

where m_i is the total number of different terms in the S_i sentence.

f_{21}: Similarity to First Sentence: This feature scores a sentence based on its similarity to the first sentence in the document. Similarity to first sentence is computed as in (6):

$$Score_{f_{21}}(S_i) = cosine_{similarty}(S_i, S_{First}) \quad (6)$$

Given two vectors of attributes, the cosine similarity is represented using a dot product. For the text summarization, the attribute vectors are the term frequency vectors of the sentences.

f_{22}: Similarity to Last Sentence: This feature scores a sentence based on its similarity to the last sentence in the document. Similarity to last sentence is computed as follows as in equation (7):

$$Score_{f_{22}}(S_i) = cosine_{similarity}(S_i, S_{Last}) \quad (7)$$

f_{23}: Similarity to Title: This feature scores a sentence based on its similarity to the title in the document. Similarity to title is computed according to (8):

$$Score_{f_{23}}(S_i) = cosine_{similarity}(S_i, Title) \quad (8)$$

f_{31}: Sentence Length: We assume that longer sentences contain more information. For a sentence S_i in a document d, the feature score is calculated as shown in (9):

$$Score_{f_{31}}(S_i) = total\ number\ of\ terms\ in\ S_i \quad (9)$$

f_{32}: Term Frequency: This feature depends on the intuition that the importance of a term for a document is directly proportional to its number of occurrences in the document [4]. In our study, each sentence is given a frequency score by summing the frequencies of the constituent words. Term frequency sentene score function can be seen in (10):

$$Score_{f_{31}}(S_i) = \sum_{k=1}^{m_i} tf(d, t_k) \quad (10)$$

where m_i is the total number of different terms in S_i and $tf(d, t_k)$ is the number of times term t_k occurs in the document d.

f_{33}: Word Sentence Score: This sentence feature is used by [6] and depends on the term frequency and inverse sentence frequency (TF_s-ISF) of t_k in S_i ($i=1,\ldots,N$) where N is the total number of sentences of the document.

The TF_S-ISF score of t_k in S_i is calculated as in equation (11):

$$TF_s - ISF(S_i, t_k) = tf(S_i, t_k) * \left[1 - \frac{\log(sf(t_k) + 1)}{\log(N + 1)}\right] \quad (11)$$

where $tf(S_i, t_k)$ is the number of times t_k occurs in S_i and $sf(t_k)$ is the number of sentences containing the term t_k.

For a sentence S_i in the document d, the f_{33} feature score is calculated according to (12):

$$Score_{f_{33}}(S_i) = 0.1 + \frac{\sum_{k=1}^{m_i} TF_s - ISF(S_i, t_k)}{HTFS} \quad (12)$$

$$|\, no\ of\ sentences\ containing\ t_k \geq \frac{1}{2} LS$$

where LS is summary length, HTFS is the highest (TF_s-ISF) summation among the sentences of the document and m_i is the total number of different terms in S_i.

f_{34}: Average Tf-Idf: This sentence feature is used by [4] and depends on the term frequency and inverse document frequency (TF_d-IDF) metric. The TF_d score for the t_k is calculated as in (13):

$$TF_d(d, t_k) = \frac{tf(d, t_k)}{\max_{i=1,\ldots,nt} tf(d, t_i)} \quad (13)$$

where $tf(d, t_k)$ is the number of times term t_k occurs in the document d and nt denotes the number of terms in d.

The TF_d-IDF score of t_k in d is calculated as in equation (14):

$$TF_d - IDF(d, t_k) = TF_d(d, t_k) * \log\left(\frac{nd}{df(c, t_k)}\right) \quad (14)$$

where nd is the total number of documents in the corpus c and the document frequency df denotes the number of documents in which the term occurs.

For a sentence S_i in the document d, the f_{34} feature score is calculated according to (15):

$$Score_{f_{34}}(S_i) = \frac{\sum_{k=1}^{m_i} TF - IDF(d, t_k)}{m_i} \quad (15)$$

where m_i is the total number of different terms in S_i.

f_{41}: Numerical data: This feature counts the number of numerical terms in a sentence according to (16). Terms that are written in numerical form sometimes convey key information about a document. The amount of information conveyed by such terms in a sentence is captured by this feature.

$$Score_{f_{41}}(S_i) = total\ number\ of\ numerical\ terms\ in\ S_i \quad (16)$$

f_{42}: Punctuation Marks: The punctuation symbols "?", "!"attract attention to the sentences they reside in. The sentences with these marks have a higher score. This sentence feature is expressed as in (17):

$$Score_{f_{42}}(S_i) = total\ number\ of\ "?"\ and\ "!"\ in\ S_i \quad (17)$$

f_{43}: Positive Words: These are words such as "in summary", "consequently", "eventually", "briefly" etc., that are commonly encountered in summaries. The score of sentences is increased whenever these keywords occur in sentences. This sentence feature is shown as in (18):

$$Score_{f_{43}}(S_i) = total\ number\ of\ positive\ words\ in\ S_i \quad (18)$$

f_{44}: Noun Phrases: This feature counts the number of nouns in a sentence according to (19). In this work, nouns are extracted by the Zemberek Software in Turkish data sets.

$$Score_{f_{44}}(S_i) = total\ number\ of\ nouns\ in\ S_i \quad (19)$$

f_{51}: LSA based features: This sentence feature is based on the study [35]. The aim of this feature is to choose sentences which are related to all important topics of the document to be summarized. After performing the singular value decomposition process on a term-sentence matrix of the document, the right singular vector matrix V^T and the diagonal matrix Σ are obtained. Then a modified latent vector space B is constructed as shown in (20):

$$B = \Sigma^2 V^T \quad (20)$$

Using the modified latent vector space B each sentence is given a sentence score by using the equation (21):

$$Score_{f_{51}}(S_i) = \sqrt{\sum_{k=1}^{r} b_{ki}^2} \quad (21)$$

where b_{ki} values are the values of the matrix B and r is the number of dimension in the new latent space. In this study r value is taken as the rank value of the matrix B.

f_{52}: Centrality: The sentence centrality is a widely used feature [6]. It is computed as a combination of the similarity, the shared friends and the shared n-grams between the sentence of interest and all the other sentences in the document, normalized by *N-1*, where *N* is the number of sentences in the document. The feature is calculated as shown in (22):

$$Score_{f_{52}}(S_i) = \frac{\sum_{j=1}^{N-1} sim(S_i,S_j) + \sum_{j=1}^{N-1} n-friends(S_i,S_j) + \sum_{j=1}^{N-1} n-grams(S_i,S_j)}{N-1} \quad (22)$$
$$| i \neq j\ sim(S_i,S_j) \succ \Phi$$

where S_j is a document sentence except S_i, N is the number of sentences in the document and Φ is the similarity threshold which is determined empirically. In this study the similarity threshold is taken as 0.16.

Combining sentence features with FAHP based system

In the proposed model, to generate a summary of a given document, all sentence feature scores are normalized to the range [0, 1]. After the normalization, the features of sentence S_k are combined by the following linear model that is shown in (23):

$$Score(S_k) = W_1 \sum_{i=1}^{2} w_{1i} f_{1i} + W_2 \sum_{i=1}^{3} w_{2i} f_{2i} + \ldots + W_5 \sum_{i=1}^{2} w_{5i} f_{5i} \quad (23)$$

where W_1 denotes the weights of the main classes and w_{ij} denotes the weights of featuresunder f_{ij} the main classes.

In this section, we describe how the weights of the features can be determined by using FAHP. In FAHP, the pairwise comparisons of the criteria, features in this study, are made by using fuzzy numbers. Since the FAHP method allows a hierarchy, the sub-features are listed under main feature classes as it can be seen from Table 1. The features are pairwisely compared with respect to the main classes and the main classes are pairwisely compared with respect to the goal (text summarization). The weight calculation for each stage is done based on the extent analysis, which was introduced by [8].

Let $X=(x_1,x_2,\ldots,x_n)$ be an object set, and $G=(g_1,g_2,\ldots,g_n)$ be a goal set. According to the method of Chang extent analysis, each object is taken and extent analysis for each goal, g_i, is performed, respectively. Therefore, m extent analysis values for each object can be obtained using the following notation [21]:

$M_{g_i}^1, M_{g_i}^2, \ldots, M_{g_i}^m$ where g_i is the goal set $(i = 1, 2, \ldots, n)$ and $M_{gi}^j (j=1,2,\ldots,m)$ are triangular fuzzy numbers. After identifying initial assumptions, the steps of Chang's extent analysis can be given in four main steps, as follows:

Step1: The value of fuzzy synthetic extent value (Se_i) with respect to the i^{th} object is represented as in (24):

$$Se_i = \sum_{j=1}^{m} M_{gi}^j \otimes \left[\sum_{i=1}^{n}\sum_{j=1}^{m} M_{gi}^j\right]^{-1} \tag{24}$$

To obtain $\sum_{j=1}^{m} M_{gi}^j$, the fuzzy addition operation of m extent analysis values for a particular matrix is performed according to (25):

$$\sum_{j=1}^{m} M_{gi}^j = \left(\sum_{j=1}^{m} l_j, \sum_{j=1}^{m} m_j, \sum_{j=1}^{m} u_j\right) \tag{25}$$

where l is the lower limit value, m is the most promising value, and u is the upper limit value. Then the fuzzy addition operation of $M_{g_i}^j$ $(j=1,2,\ldots,m)$ values such that

$$\left[\sum_{i=1}^{n}\sum_{j=1}^{m} M_{gi}^j\right] = \left(\sum_{i=1}^{n} l_i, \sum_{i=1}^{n} m_i, \sum_{i=1}^{n} u_i\right)$$

are performed to obtain

$$\left[\sum_{i=1}^{n}\sum_{j=1}^{m} M_{gi}^j\right]^{-1}.$$

At the end of the $Step_1$, the inverse of the determined vector can be expressed as in (26):

$$\left[\sum_{i=1}^{n}\sum_{j=1}^{m} M_{gi}^j\right]^{-1} = \left(\frac{1}{\sum_{i=1}^{n} u_i}, \frac{1}{\sum_{i=1}^{n} m_i}, \frac{1}{\sum_{i=1}^{n} l_i}\right) \tag{26}$$

Step2: As $M_1=(l_1,m_1,u_1)$ and $M_2=(l_2,m_2,u_2)$ are two triangular fuzzy numbers, the degree of possibility of $M_2=(l_2,m_2,u_2) \geq M_1=(l_1,m_1,u_1)$ is defined as $V(M_2 \geq M_1) = \sup_{y \geq x}\left[\min(\mu_{M_1}(x), \mu_{M_2}(y))\right]$ and it can be represented as in (27):

$$V(M_2 \geq M_1) = hgt(M_1 \cap M_2) = \mu_{M_2}(d) = \begin{cases} 1, & if\ m_2 \geq m_1 \\ 0, & if\ l_1 \geq u_2 \\ \dfrac{(l_1-u_2)}{(m_2-u_2)-(m_1-l_1)}, & otherwise \end{cases} \tag{27}$$

where d is the ordinate of the highest intersection point between μ_{M_1} and μ_{M_2}. In order to compare M_1 and M_2 we need both the values of $V(M_2 \geq M_1)$ and $V(M_1 \geq M_2)$.

Step3: The degree possibility for a convex fuzzy number to be greater than k convex fuzzy M_i $(i=1,2,\ldots,k)$ can be defined by: $V(M \geq M_1, M_2, \ldots, M_k) = V[(M \geq M_1)\ and\ (M \geq M_2) \ldots and\ (M \geq M_k)] = min\, V(M \geq M_i),\ i = 1,2,\ldots,k$

Assume that $d'(A_i) = \min V(Se_i \geq Se_k)$ for $k = 1,2,\ldots,n$ and $k \neq i$. Then the weight vector is given by $W' = (d'(A_1), d'(A_2), \ldots, d'(A_n))^T$ where $A_i = (1,2,\ldots,n)$ are n elements.

Step4: Via normalization, the normalized weight vectors are $W = (d(A_1), d(A_2), \ldots, d(A_n))^T$ where W is a non-fuzzy number and d is the coordinate of highest intersection point D between μ_{M_1} and μ_{M_2} (see Figure 1).

Figure 1

The relation between M_1 and M_2 [8]

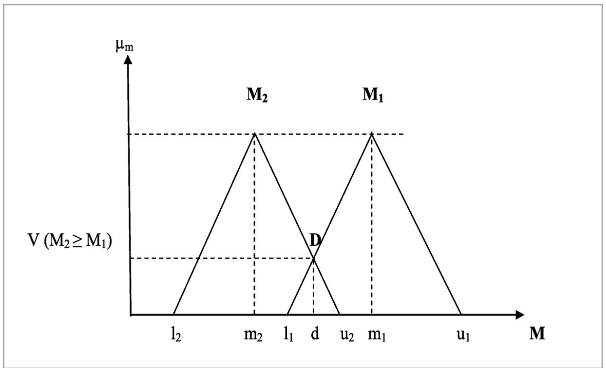


Table 2

Linguistic scale and corresponding Triangular Fuzzy Scale

Linguistic Scale	Triangular Fuzzy Scale	Triangular Fuzzy Reciprocal Scale
Just Equal	(1, 1, 1)	(1, 1, 1)
	(1/2, 3/4, 1)	(1, 4/3, 2)
Weakly Important	(2/3, 1, 3/2)	(2/3, 1, 3/2)
	(1, 3/2, 2)	(1/2, 2/3, 1)
Strongly More Important	(3/2,2,5/2)	(2/5,1/2,2/3)
	(2, 5/2, 3)	(1/3, 2/5, 1/2)
Very strong more important	(5/2,3,7/2)	(2/7,1/3,2/5)
	(3, 7/2, 4)	(1/4, 2/7, 1/3)
Absolutely more important	(7/2,4,9/2)	(2/9,1/4,2/7)

Application of the proposed FAHP based model

In this methodology, as a first step, we have analyzed each normalized sentence feature in a hierarchal structure as shown in Table 1. To create pairwise comparison matrix, fuzzy linguistic scale is used which is given in Table 2. Different scales can be found in the literature as in the study [1].

Then, we ask three linguistics experts to construct pairwise comparison matrices that indicate how many times more important one feature is with respect to another one. Table 3 shows the aggregated fuzzy pairwise comparisons of the three experts for the main classes with respect to the goal (text summarization).

According to the FAHP method, firstly synthetic values must be calculated. From Table 3, synthetic values with respect to main goal are calculated like in equation (24). These fuzzy values are compared by using equation (27) Then priority weights [W' = (1, 0.976, 0.988, 0.835, 0.936)] are calculated as described in $Step_3$ in Section 4. After the normalization of these values priority weights with respect to main goal are calculated as W = (0.2111, 0.2061, 0.2086, 0.1764, 0.1978).

Tables 3-8 show the aggregated fuzzy pairwise comparisons of the features under the main feature classes respectively. Table 4 shows the aggregated fuzzy pairwise comparisons of the three experts for the Sentence Location (f_{11}) and Distributional Features (f_{12}) with respect to the main feature Location Knowledge (F_1). Table 5, 6, 7, and 8 show the aggregated fuzzy pairwise comparisons of the three experts for the features under main classes F_2, F_3, F_4, F_5, respectively.

After forming fuzzy pair-wise comparison matrices, weights of all features under the main groups are determined by the same steps of FAHP. Both the hierarchy and the weights of all of the features are shown in Figure 2.

Table 3

Aggregated pairwise comparison matrix of the main features with respect to the goal

	F_1	F_2	F_3	F_4	F_5
F_1	(1, 1, 1)	(2/3, 1, 3/2)	(1/2, 3/4, 1)	(3/2, 2, 5/2)	(1/2, 3/4, 1)
F_2	(2/3, 1, 3/2)	(1, 1, 1)	(1, 4/3, 2)	(2/3, 1, 3/2)	(2/3, 1, 3/2)
F_3	(1, 4/3, 2)	(1/2, 3/4, 1)	(1, 1, 1)	(2/3, 1, 3/2)	(1, 4/3, 2)
F_4	(2/5, 1/2, 2/3)	(2/3, 1, 3/2)	(2/3, 1, 3/2)	(1, 1, 1)	(2/3, 1, 3/2)
F_5	(1, 4/3, 2)	(2/3, 1, 3/2)	(1/2, 3/4, 1)	(2/3, 1, 3/2)	(1, 1, 1)

Table 4

Aggregated pairwise comparison matrix for the features under F1

	f_{11}	f_{12}
f_{11}	(1, 1, 1)	(1/2, 3/4, 1)
f_{12}	(1, 4/3, 2)	(1, 1, 1)

Table 5

Aggregated pairwise comparison matrix for the features under F2

	f_{21}	f_{22}	f_{23}
f_{21}	(1, 1, 1)	(2/3, 1, 3/2)	(1, 4/3, 2)
f_{22}	(2/3, 1, 3/2)	(1, 1, 1)	(1, 4/3, 2)
f_{23}	(1/2, 3/4, 1)	(1/2, 3/4, 1)	(1, 1, 1)

Table 6

Aggregated pairwise comparison matrix for the features under F3

	f_{31}	f_{32}	f_{33}	f_{34}
f_{31}	(1, 1, 1)	(2/3, 1, 3/2)	(1, 4/3, 2)	(1/2, 3/4, 1)
f_{32}	(2/3, 1, 3/2)	(1, 1, 1)	(2/3, 1, 3/2)	(1, 4/3, 2)
f_{33}	(1/2, 3/4, 1)	(2/3, 1, 3/2)	(1, 1, 1)	(2/3, 1, 3/2)
f_{34}	(1, 4/3, 2)	(1/2, 3/4, 1)	(2/3, 1, 3/2)	(1, 1, 1)

Table 7

Aggregated pairwise comparison matrix for the features under F4

	f_{41}	f_{42}	f_{43}	f_{44}
f_{41}	(1, 1, 1)	(1/2, 3/4, 1)	(2/3, 1, 3/2)	(2/5, 1/2, 2/3)
f_{42}	(1, 4/3, 2)	(1, 1, 1)	(2/3, 1, 3/2)	(2/3, 1, 3/2)
f_{43}	(2/3, 1, 3/2)	(2/3, 1, 3/2)	(1, 1, 1)	(1, 4/3, 2)
f_{44}	(3/2, 2, 5/2)	(2/3, 1, 3/2)	(1/2, 3/4, 1)	(1, 1, 1)

Table 8

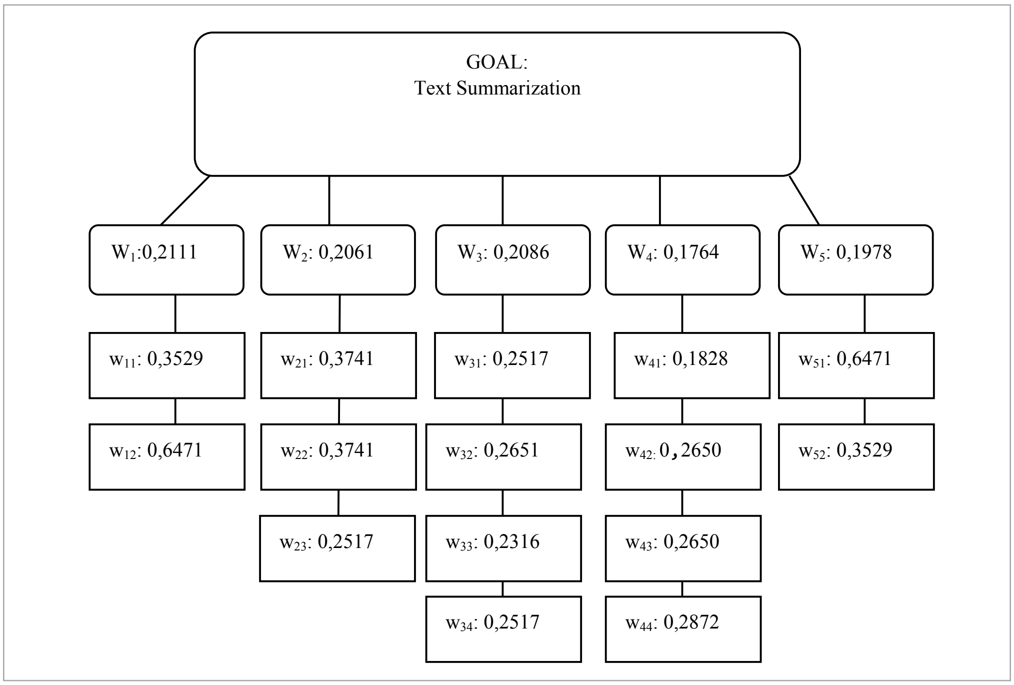
Aggregated pairwise comparison matrix for the features under F5

	f_{51}	f_{52}
f_{51}	(1, 1, 1)	(1, 4/3, 2)
f_{52}	(1/2, 3/4, 1)	(1, 1, 1)

It can be seen from Figure 2 that the highest weight among the main classes belongs to the F_1, while the weights of F_2 and F_3 are very close to each other. Under the class F_1, the weight of f_{12} is higher than f_{11}. Under the class F_2, the weights of f_{21} and f_{22} are equal and higher than the weight of f_{23}. Under the classes F_3, F_4 and F_5, the weights of f_{32}, f_{44} and f_{51} are the highest ones.

Figure 2

Hierarchy of the features affecting text summarization

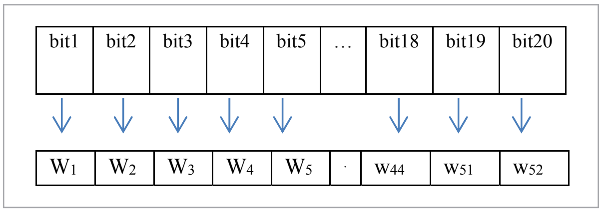


Combining sentence features with Genetic Algorithm

GA is the most widely used approach for computational evolution. It is inspired by the various techniques of natural evolution like selection, crossover and mutation. GA begins with a set of solutions (chromosomes) called population. In this phase, the most important thing is how to represent the problem using chromosomes and how to build coding scheme for genes of a chromosome. There are two common coding schemes for gene positions (bits) of the chromosomes; real-valued or binary encoding [34]. In this study the real-valued encoding where each chromosome is coded as a vector of real numbers between 0-1 is used. Since there are 15 sentence features under five different classes, our chromosome structures contain 20 (15+5) gene positions (see Figure 3).

Figure 3

Chromosome structure for features



The first five bits refer to the weights of the main classes and the others refer to the features under those classes. After encoding the parameters, the population of 100 chromosomes is randomly generated at the beginning. The fitness of a chromosome in GA is the value returned by the fitness evaluation function. This evaluation function measures the quality of chromosomes to solve a problem. In this study, the document sentences are scored using (Eq. 22) and ranked in a descending order according to their scores. A set of the highest scoring sentences are extracted as a document summary based on the compression rate. In this study, we used a 35% compression rate as summary length. Then, the created summary is used as input for the fitness function (28) which obtains the best average recall score generated:

$$fitness = \frac{|S \cap T|}{|T|} \qquad (28)$$

where T is the reference summary and S is the system generated summary. In order to implement GA, java genetic algorithm package (JGAP) is used. We used the value of 80% as crossover probability and 10% as mutation probability.

Data corpus and the evaluation dataset

In order to see the performance of the proposed system, Turkish datasets are used. Turkish data sets contain two different sets of Turkish documents. The first set (Turkish130) involves 130 documents related to different areas and a human-generated extractive summary corpus which is created with 35% summarization ratio. Table 9 shows the attributes of the Turkish130 data corpus.

Table 9

Attributes of the Turkish130 data corpus

Attributes of the data corpus	Values
Number of docs	130
Total number of sentences	2487
Min sentences/doc	9
Max sentences/doc	63

The second set (Turkish20) contains 20 documents and 30 extractive summary corpora which are prepared by 30 different assessors (15 male, 15 female). The aim under the use of Turkish20 is to show the stability in FAHP based system's result. The documents in the first set are longer than the documents in the second set. Table 10 shows attributes of the Turkish20 data corpus.

Table 10

Attributes of the Turkish20 data corpus

Attributes of the data corpus	Values
Number of docs	20
Total number of sentences	201
Min sentences/doc	7
Max sentences/doc	10

Experimental results

Performance analysis is conducted on each individual sentence feature, FAHP based combined system, and GA based combined system using the prepared Turkish data sets. While analyzing the performance, human generated summaries are compared with the automatically generated summaries. The precision (P), recall (R), and F-measure (F) metrics that enable the evaluation of sentence coverage among manually and automatically generated summaries are chosen for evaluation results. Assuming that T is the reference summary and S is the system generated summary, the measurements P, R and F are defined as follows:

$$P = \frac{|S \cap T|}{|S|}, \quad R = \frac{|S \cap T|}{|T|}, \quad F = \frac{2PR}{R+P} \tag{29}$$

Table 11

Performance results of each feature, the proposed FAHP based system and GA based system of Turkish130 dataset

The main Classes	Sentence features	$Assessor_1$
F_1	f_{11}	0,423
	f_{12}	0,553
F_2	f_{21}	0,441
	f_{22}	0,371
	f_{23}	0,394
F_3	f_{31}	0,539
	f_{32}	0,451
	f_{33}	0,541
	f_{34}	0,445
F_4	f_{41}	0,182
	f_{42}	0,002
	f_{43}	0,103
	f_{44}	0,532
F_5	f_{51}	0,511
	f_{52}	0,508
Combining features by FAHP		0,562
Combining features by GA		0,565

Table 11 summarizes the effects of each feature, the proposed FAHP based system and GA based system on Turkish130 dataset.

Considering the ordering of the features with respect to their performances, one can say that the assessor mostly gives more importance to the distributional feature (f_{12}), word sentence score (f_{33}), and sentence length (f_{31}). It is seen from the table that the assessor gives less importance to numerical data (f_{41}), positive key words (f_{43}) and punctuation marks (f_{42}). The results in Table 11 show that exploiting all features by combining them resulted in a better performance (0,562 and 0,565) than exploiting each feature individually. GA based combining method requires training and testing which is time-consuming and depends on the availability of the labelled document dataset. For the training and testing phases, leave-one-out cross-validation (LOOCV) is used. In this method the system would initially be trained on 129 of the documents and tested on the reaming one. This procedure is repeated for a further 129 times, omitting a different document at each time from the training data, resetting the sentence score weights to random values, retraining and then testing on the omitted document. By this way, the performance of the summarization system can be evaluated using every available document as though it was previously unseen by the test data. From Table 11, it is seen that FAHP based combining method presents very similar results with GA. Since many experiments are conducted, there is a standard deviation value for the GA-based combining method. The standard deviation value for our study is 0,1749. FAHP incorporates expert knowledge and opinion in the problem and once the weights are determined by the expert, these weights could be used on all other similar datasets. Not only FAHP but also GA generates domain-dependent feature weights, which means that if the dataset is changed, then the produced weights have to be recomputed.

Table 12 shows the F-measure values of each feature under the main classes and the effect of the proposed FAHP based system on Turkish20 data set. The last row of this table indicates the average performances of each feature and FAHP based system with respect to 30 different assessors (15 male, 15 female). In this table, Ma_i denotes the i^{th} male assessor and Fe_i denotes the i^{th} female assessor.

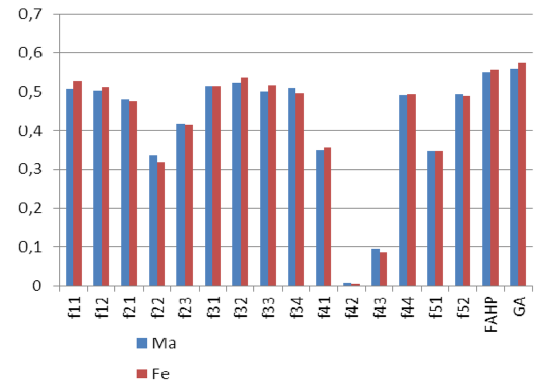
When the last row of Table 12 is sorted, it is seen

Table 12

Performance results of each feature, the proposed FAHP based system and GA based system of Turkish20 dataset

	F_1		F_2			F_3				F_4				F_5		FAHP	GA
	f_{11}	f_{12}	f_{21}	f_{22}	f_{23}	f_{31}	f_{32}	f_{33}	f_{34}	f_{41}	f_{42}	f_{43}	f_{44}	f_{51}	f_{52}		
Ma_1	0,480	0,580	0,530	0,291	0,468	0,584	0,493	0,572	0,480	0,418	0,000	0,079	0,599	0,410	0,545	0,597	0,576
Ma_2	0,623	0,538	0,551	0,289	0,488	0,566	0,603	0,558	0,569	0,438	0,000	0,096	0,510	0,356	0,524	0,643	0,680
Ma_3	0,603	0,518	0,493	0,320	0,393	0,530	0,582	0,522	0,561	0,373	0,013	0,092	0,474	0,303	0,503	0,597	0,586
Ma_4	0,535	0,538	0,441	0,406	0,391	0,553	0,585	0,512	0,568	0,281	0,013	0,133	0,495	0,341	0,514	0,576	0,564
Ma_5	0,493	0,551	0,516	0,275	0,374	0,566	0,530	0,583	0,493	0,395	0,013	0,117	0,560	0,353	0,537	0,618	0,620
Ma_6	0,537	0,470	0,534	0,349	0,472	0,424	0,553	0,428	0,541	0,320	0,013	0,121	0,410	0,372	0,423	0,503	0,531
Ma_7	0,362	0,601	0,462	0,377	0,412	0,638	0,445	0,576	0,362	0,395	0,000	0,117	0,558	0,459	0,643	0,593	0,649
Ma_8	0,468	0,480	0,349	0,288	0,353	0,474	0,443	0,487	0,438	0,316	0,000	0,096	0,473	0,374	0,458	0,476	0,497
Ma_9	0,509	0,516	0,516	0,306	0,408	0,523	0,559	0,502	0,518	0,345	0,013	0,108	0,508	0,298	0,518	0,574	0,603
Ma_{10}	0,563	0,486	0,476	0,308	0,401	0,482	0,497	0,478	0,576	0,343	0,000	0,038	0,484	0,288	0,484	0,553	0,565
Ma_{11}	0,476	0,378	0,491	0,406	0,449	0,381	0,484	0,343	0,484	0,308	0,013	0,083	0,325	0,306	0,356	0,445	0,403
Ma_{12}	0,468	0,463	0,484	0,445	0,405	0,474	0,488	0,453	0,522	0,299	0,013	0,083	0,491	0,318	0,460	0,53	0,534
Ma_{13}	0,470	0,447	0,399	0,350	0,374	0,458	0,487	0,408	0,491	0,249	0,013	0,121	0,389	0,324	0,416	0,447	0,447
Ma_{14}	0,528	0,563	0,503	0,288	0,424	0,578	0,591	0,583	0,495	0,408	0,000	0,079	0,568	0,370	0,562	0,63	0,612
Ma_{15}	0,505	0,433	0,470	0,358	0,445	0,474	0,509	0,474	0,534	0,336	0,013	0,079	0,499	0,328	0,458	0,47	0,489
Fe_1	0,608	0,447	0,547	0,289	0,476	0,430	0,520	0,447	0,558	0,306	0,013	0,079	0,416	0,255	0,403	0,534	0,534
Fe_2	0,547	0,580	0,451	0,273	0,388	0,558	0,563	0,574	0,522	0,390	0,000	0,067	0,527	0,343	0,556	0,597	0,648
Fe_3	0,648	0,412	0,495	0,277	0,412	0,387	0,594	0,441	0,623	0,268	0,013	0,096	0,389	0,249	0,360	0,533	0,615
Fe_4	0,576	0,512	0,499	0,298	0,453	0,539	0,530	0,518	0,509	0,295	0,013	0,071	0,529	0,352	0,535	0,616	0,637
Fe_5	0,451	0,512	0,395	0,333	0,420	0,553	0,418	0,508	0,376	0,424	0,000	0,083	0,493	0,327	0,518	0,512	0,608
Fe_6	0,537	0,488	0,499	0,339	0,399	0,441	0,583	0,483	0,528	0,314	0,000	0,079	0,498	0,316	0,424	0,518	0,528
Fe_7	0,368	0,512	0,402	0,441	0,293	0,520	0,368	0,495	0,327	0,349	0,000	0,083	0,491	0,437	0,537	0,487	0,484
Fe_8	0,474	0,553	0,458	0,277	0,391	0,648	0,541	0,573	0,491	0,403	0,013	0,104	0,621	0,402	0,643	0,562	0,628
Fe_9	0,526	0,529	0,523	0,273	0,457	0,522	0,551	0,551	0,443	0,388	0,000	0,067	0,491	0,380	0,483	0,588	0,638
Fe_{10}	0,433	0,474	0,428	0,314	0,383	0,556	0,441	0,502	0,387	0,420	0,000	0,096	0,513	0,406	0,514	0,512	0,512
Fe_{11}	0,669	0,545	0,508	0,254	0,483	0,548	0,703	0,589	0,632	0,333	0,013	0,104	0,533	0,318	0,518	0,649	0,708
Fe_{12}	0,590	0,516	0,470	0,343	0,370	0,474	0,594	0,508	0,603	0,422	0,000	0,079	0,448	0,324	0,423	0,562	0,513
Fe_{13}	0,426	0,505	0,463	0,420	0,384	0,474	0,522	0,466	0,455	0,301	0,013	0,108	0,381	0,373	0,435	0,472	0,422
Fe_{14}	0,573	0,561	0,540	0,277	0,444	0,499	0,569	0,545	0,544	0,323	0,000	0,079	0,510	0,341	0,448	0,578	0,573
Fe_{15}	0,480	0,538	0,469	0,362	0,473	0,563	0,576	0,551	0,443	0,418	0,013	0,104	0,558	0,368	0,537	0,601	0,580
average	0,517	0,508	0,479	0,327	0,416	0,514	0,531	0,507	0,502	0,353	0,007	0,091	0,491	0,346	0,491	**0,552**	**0,566**

Figure 4
Comparison of the performance results with respect to genders



f_{11}	f_{12}	f_{21}	f_{22}	f_{23}	f_{31}	f_{32}	f_{33}	f_{34}
0,508 0,527	0,504 0,512	0,481 0,476	0,337 0,318	0,417 0,415	0,514 0,514	0,523 0,538	0,499 0,517	0,509 0,496

f_{41}	f_{42}	f_{43}	f_{44}	f_{51}	f_{52}	**FAHP**	**GA**
0,348 0,357	0,008 0,006	0,096 0,087	0,490 0,493	0,347 0,346	0,493 0,489	**0,550** **0,555**	**0,557** **0,575**

that the assessors tend to use f_{32}-"Term Frequency"; f_{11}-"Sentence Location"; f_{31}-"Sentence Length"; f_{12}-"Distributional Features"; f_{33}-"Word Sentence Score"; f_{34}-"Average Tf-Idf"; f_{44}-"Noun Phrase"; f_{51}-"LSA based Feature"; f_{21}-"Similarity to First Sentence"; f_{23}-"Similarity to Title Sentence"; f_{41}-"Numerical Data"; f_{52}-"Centrality"; f_{22}-"Similarity to Last Sentence"; f_{43}-"Positive key words" and f_{42}-"Punctuation Marks" features respectively. From the results, it can be said that combining sentence features via FAHP gives better performance results (0,552) than each individual features. GA based sentence scoring method achieves a slightly better performance but as mentioned before the supervised structure of this method involves resetting the sentence score weights, training and testing and then scoring the sentences many times- for each assessor, 20 times in this example. However, FAHP based method has a very close performance value and advantageously can use the pre-determined weights – it does not have to reset them.

When the average performance of each feature is considered separately with respect to summaries created by male and female assessors, it is seen from Figure 4 that they draw almost the same pattern. According to this figure, one can say that male and female assessors think in very similar way while they are creating the summary documents. It is also seen that FAHP based combining method shows better performance result than all individual sentence features. It can also be depicted from Figure 4 that FAHP performs as high as GA.

Data corpus and the evaluation dataset

This paper proposes a novel text summarization system that combines various sentence features to enhance summarization results. For transforming the features' scores into an overall score function, FAHP is employed. FAHP determines the weights of the criteria and sub-criteria (features in this study) in order to be used in the calculation of the overall score function. FAHP method has the advantages of allowing the human involvement in the weight determination stage, considering both quantitative and qualitative criteria, eliminating a training phase and dealing with situations characterized by imprecision due to subjective and qualitative evaluation. To the best of our knowledge, this study is the first to employ FAHP in the area of determining the weights of the features affecting the text summarization process. Moreover, it divides the features into a hierarchy, which allows to calculate both the weights of the main criteria and sub-criteria (i.e. features). These weights are then used in an additive weighting method and a linear model is proposed in order to obtain an overall score function.

To carry out the experiments for the comparison of the proposed system against systems employing only structural or only semantic features, two different Turkish corpora have been constructed. Unlike automatic combining methods, FAHP generates general feature weights depending on the experts' judgements without a training phase. Although the performance improvements are not much significant, which is usually the case in most of research in this area, the obtained improvements show that FAHP can be used for deciding the contribution level of each feature for important sentence selection phase. Furthermore, the computational results are compared with the results of the GA based method in order to show the usefulness of the proposed method.

In brief, this paper concludes with 4 main contributions to the literature:

1 It is more useful to combine the weights of the features instead of using them individually,

2 FAHP is an effective way for text summarization,

3 FAHP contributes to the text summarization research area by handling uncertainty and allowing the human involvement in the weight determination stage,

4 The performance results of the FAHP method are approximately the same as GA, and advantageously FAHP method saves time compared to the GA algorithm. This is due to the fact that FAHP method determines the weights of the features once it asks to the experts at a time, whereas GA needs many experiments to decide on these weights.

Although this study has many contributions to the literature, as a further research; the comparison of the proposed method and the meta-heuristic based automatic methods can be conducted. The performance results of these experiments will show a pathway for both practitioners and researchers who study on text summarization area. Another future research topic can be increasing the number of features and detecting the most useful ones among them. Moreover, since this study has the limitation of using only Turkish corpora, it can be extended by adding datasets in other languages.

References

1. M. G. Abdel-Kader, D. Dugdale. Evaluating investments in advanced manufacturing technology: A fuzzy set theory approach. The British Accounting Review, 2001, 33(4), 455-489. https://doi.org/10.1006/bare.2001.0177
2. E. Adamcsek. The Analytic Hierarchy Process and its Generalizations. Unpublished doctoral dissertation, Thesis, Eotvos Lorand University, 2008.
3. R. M. Alguliev, R. M. Aliguliyev, M. S. Hajirahimova, C. A. Mehdiyev. MCMR: Maximum coverage and minimum redundant text summarization model. Expert Systems with Applications, 2011, 38(12), 14514-14522. https://doi.org/10.1016/j.eswa.2011.05.033
4. M. Berker, T. Güngör. Using Genetic Algorithms with Lexical Chains for Automatic Text Summarization. 4th International Conference on Agents and Artificial Intelligence, Vilamoura, Portugal, 2012.
5. H. Bhandari, M. Shimbo, T. Ito, Y. Matsumoto. Generic Text Summarization Using Probabilistic Latent Semantic Indexing. The Third International Joint Conference on Natural Language Processing, 2008.
6. M. S. Binwahlan, N. Salim, L. Suanmali. Fuzzy Swarm Diversity Hybrid Model for Text Summarization. Information Processing and Management, 2010, 46(5), 571-588. https://doi.org/10.1016/j.ipm.2010.03.004
7. F. T. S. Chan, N. Kumar. Global supplier development considering risk factors using fuzzy extended AHP-based approach. Omega, 2007, 35(4), 417–431. https://doi.org/10.1016/j.omega.2005.08.004
8. D. Y. Chang. Applications of the extent analysis method on Fuzzy Analytical Hierarchy Process. European Journal of Operational Research, 1996, 95 (3), 649–655. https://doi.org/10.1016/0377-2217(95)00300-2
9. V. Y. C. Chen, H. P. Lien, C. H. Liu, J. J. H. Liou, G. H. Tzeng, L. S. Yang. Fuzzy MCDM approach for selecting the best environment-watershed plan. Applied Soft Computing, 2011, 11, 265–275. https://doi.org/10.1016/j.asoc.2009.11.017
10. Ö. Coşgun, Y. Ekinci, S. Yanık. Fuzzy rule-based demand forecasting for dynamic pricing of a maritime company. Knowledge-Based Systems, 2014, 70, 88-96. https://doi.org/10.1016/j.knosys.2014.04.015
11. H. P. Edmundson. New methods in automatic extracting. Journal of the Association for Computing Machinery, 1969, 16(2), 264–285. https://doi.org/10.1145/321510.321519
12. E. Filatova, V. Hatzivassiloglou. A formal model for information selection in multi-sentence text extraction. Proceedings of the 20th International Conference on

Computational Linguistics, 2004, 397-403. https://doi.org/10.3115/1220355.1220412

13. M. A. Fattah, F. Ren. GA. MR, FFNN, PNN and GMM based models for automatic text summarization. Computer Speech and Language, 2009, 23(1), 126-144. https://doi.org/10.1016/j.csl.2008.04.002
14. Y. Gong, X. Liu. Generic Text Summarization Using Relevance Measure and Latent Semantic Analysis. In: The Proceedings of ACM SIGIR, New Orleans, Louisiana, United States, 2001, 19-25. https://doi.org/10.1145/383952.383955
15. A. Güran, N. G. Bayazıt, M. Z. Gurbuz. Efficient feature integration with Wikipedia-based semantic feature extraction for Turkish text summarization. Turkish Journal of Electrical Engineering and Computer Sciences, 2013, 21(5), 1411-1425. https://doi.org/10.3906/elk-1201-15
16. A. Güran, N. G. Bayazıt, E. Bekar. Automatic summarization of Turkish documents using non-negative matrix factorization. Innovations in Intelligent Systems and Applications, 2011, 480-484. https://doi.org/10.1109/inista.2011.5946121
17. A. Güran, N. G. Bayazıt. A New Preprocessing Phase for LSA-Based Turkish Text Summarization. In: Recent Advances in Computer Science and Information Engineering, Springer, Berlin-Heidelberg, 2012, 305-310. https://doi.org/10.1007/978-3-642-25781-0_46
18. A. Güran, E. Bekar, S. Akyokus. A comparison of feature and semantic-based summarization algorithms for Turkish. In: International Symposium on Innovations in Intelligent Systems and Applications, 2010, 21-24.
19. R. A. G Hernandez, Y. Ledeneva. Word Sequence Models for Single Text Summarization. In: Proceedings of Second International conference on Advances in Computer-Human Interactions, Mexico, 2009, 44-48.
20. B. J. Johansson. Achieving high commitment in and of sourcing decisions–a decision-making paradox. Information Technology and Control, 2015, 35(3), 313-322.
21. C. Kahraman, Z. Ulukan, E. Tolga. A fuzzy weighted evaluation method using objective and subjective measures. Proceedings of the International ICSC Symposium on Engineering of Intelligent Systems, Spain, 1998.
22. W. Karwowski, R. W. Evans. Fuzzy concepts in production management research: a review. International Journal of Production Research, 1986, 24(1), 129-147. https://doi.org/10.1080/00207548608919718
23. F. Kyoomarsi, H. Khosravi, E. Eslami, P. K Dehkordy. Optimizing Machine Learning Approach Based on Fuzzy Logic in Text Summarization. International Journal of Hybrid Information Technology, 2009, 2(2), 105-116.
24. J. H. Lee, S. Park, C. M. Ahn, D. Kim. Automatic generic document summarization based on non-negative matrix factorization. Information Processing and Management, 2009, 45(1), 20–34. https://doi.org/10.1016/j.ipm.2008.06.002
25. H. P. Luhn. The automatic creation of literature abstracts. IBM Journal of Research Development, 1958, 2, 159–165. https://doi.org/10.1147/rd.22.0159
26. R. McDonald. A study of global inference algorithms in multi-document summarization. European Conference on Information Retrieval. Springer, Berlin-Heidelberg, 2007, 557-564. https://doi.org/10.1007/978-3-540-71496-5_51
27. Y. K. Meena, D. Gopalani. Evolutionary Algorithms for Extractive Automatic Text Summarization. Procedia Computer Science, 2005, 48, 244-249. https://doi.org/10.1016/j.procs.2015.04.177
28. G. Murray, S. Renals, J. Carletta. Extractive summarization of meeting recordings. In: Proceedings of the 9th European Conference on Speech Communication and Technology, Lisbon, Portugal, 2005.
29. F. Ozgur Catak, S. Karabas, S. Yildirim. Fuzzy Analytic Hierarchy Based Database Management Systems Selection. In: Turkish National Identity Card Management Project. International Journal of Information Sciences and Techniques, 2011, 2(4), 29-38. https://doi.org/10.5121/ijist.2012.2403
30. T. Pardo, L. Rino, M. Nunes. GistSumm: A summarization tool based on a new extractive method. In: 6th Workshop on Computational Processing of the Portuguese Language, São Carlos, Brazil, 2003, 210-218. https://doi.org/10.1007/3-540-45011-4_34
31. R. Piotrowski, A. Skiba. Nonlinear Fuzzy Control System for Dissolved Oxygen with Aeration System in Sequencing Batch Reactor. Information Technology and Control, 2015, 44(2), 182-195. https://doi.org/10.5755/j01.itc.44.2.7784
32. Y. Quyang, W. Li, Q. Lu, R. Zhang. A study on Position Information in Document Summarization. In: Proceedings of Coling, Poster Volume, Beijing, 2010, 919-927.
33. R. Radev, S. Blair-Goldensohn, Z. Zhang. Experiments in Single and Multi-Docuement Summarization using MEAD. In: First Document Understanding Conference, New Orleans, LA, 2001.
34. Y. Rahmat-Samii, E. Michielssen. Electromagnetic Optimization by Genetic Algorithms. Microwave Journal, 1999, 42(11), 232-232.

35. J. Steinberger. Text Summarization within the LSA Framework. PhD Thesis, University of West Bohemia in Pilsen, Czech Republic, 2007.

36. S. M. Tofighy, R. G. Raj, H. H. S. Javadi. AHP techniques for Persian text summarization. Malaysian Journal of Computer Science, 2013, 26(1), 1-8.

37. E. Uzundere, E. Dedja, B. Diri, M. F. Amasyalı. Türkçe Haber Metinleri İçin Otomatik Özetleme. Akıllı Sistemlerde Yenilikler ve Uygulamaları Sempozyumu, Isparta, Türkiye, 2008, 1-3.

38. R. Witte, R. Krestel, S. Bergler. Generating Update Summaries for DUC 2007. In: The Document Understanding Workshop, Rochester, NewYork, USA, 26-27, 2007.

39. X. B. Xue, Z. H. Zhou. Distributional features for text categorization. IEEE Transactions on Knowledge and Data Engineering, 2009, 21(3), 428-442. https://doi.org/10.1109/TKDE.2008.166

40. J. Y. Yeh, H. R. Ke, W. P. Yang, I. H. Meng, Text Summarization using a Trainable Summarizer and Latent Semantic Analysis. Information Processing and Management, 2005, 41, 75–95. https://doi.org/10.1016/j.ipm.2004.04.003

41. L. A. Zadeh. Fuzzy Sets. Information and Control, 1965, 8, 338-353. https://doi.org/10.1016/S0019-9958(65)90241-X

42. Zemberek. http://code.google.com/p/zemberek/. Accessed on March 1, 2008.

43. G. K. Zipf. Human Behavior and the Principle of Least Effort. Ravenio Books, 2016.

Summary / Santrauka

This study proposes a novel additive Fuzzy Analytical Hierarchy Process (FAHP) based sentence score function for Automatic Text Summarization (ATS), which is a method to handle growing amounts of textual data. ATS aims to reduce the size of a text while covering the important points in the text. For this aim, this study uses some sentence features, combines these features by an additive score function using some specific weights and produces a sentence score function. The weights of the features are determined by FAHP – specifically Fuzzy Extend Analysis (FEA), which allows the human involvement in the process, uses pair-wise comparisons, addresses uncertainty and allows a hierarchy composed of main features and sub-features. The sentences are ranked according to their score function values and the highest scored sentences are extracted to create summary documents. Performance evaluation is based on the sentence coverage among the summaries generated by human and the proposed method. In order to see the performance of the proposed system, two different Turkish datasets are used and as a performance measure, the F-measure is used. The proposed method is compared with a heuristic algorithm, namely Genetic Algorithm (GA). Resulting performance improvements show that the proposed model will be useful for both researchers and practitioners working in this research area.

Ši studija siūlo naują papildomą miglotųjų aibių analitinio hierarchijos proceso (angl. Fuzzy Analytical Hierarchy Process (FAHP)) funkciją automatinio teksto santraukos generavimo metodui (angl. Automatic Text Summarization (ATS)), kuris padeda susidoroti su augančiais tekstinių duomenų kiekiais. ATS siekia sumažinti teksto apimtį, išryškindamas svarbiausius jo elementus. Šis tyrimas naudoja kai kurias sakinio ypatybes, kombinuoja jas papildoma rezultato funkcija, naudojant tam tikrus specifinius svorius, ir pateikia sakinio rezultato funkciją. Ypatybių svoriai yra nustatomi miglotosios išplėstinės analizės (angl. Fuzzy Extended Analysis (FEA)), kuri leidžia žmogaus dalyvavimą procese, naudoja porinius palyginimus, sprendžia miglotumo problemą, leidžia sudaryti svarbiausių ypatybių ir subypatybių hierarchiją. Sakiniai išrikiuojami pagal jų rezultato funkcijos vertes, aukščiausi atsidūrę sakiniai išimami, iš jų sukuriama dokumento santrauka. Atlikimo kokybės įvertinimas pagrįstas sakinių panašumu tarp santraukų, sukurtų žmogaus ir siūlomo metodo principu. Tam, kad būtų galima įvertini siūlomo metodo veiksmingumą, pasirinkti du skirtingi turkiški duomenų rinkiniai ir F įvertis. Pasiūlytas metodas lyginamas su euristiniu algoritmu, konkrečiau – Genetiniu algoritmu (angl. Genetic Algorithm (GA)). Veiksmingumo rezultatai rodo, kad siūlomas modelis bus naudingas tiek tyrėjams, tiek praktikams, kurie dirba šioje srityje.