EXTRACTION-BASED TEXT SUMMARIZATION USING FUZZY ANALYSIS

F. KYOOMARSI, H. KHOSRAVI, E. ESLAMI AND M. DAVOUDI

Abstract. Due to the explosive growth of the world-wide web, automatic text summarization has become an essential tool for web users. In this paper we present a novel approach for creating text summaries. Using fuzzy logic and word-net, our model extracts the most relevant sentences from an original document. The approach utilizes fuzzy measures and inference on the extracted textual information from the document to find the most significant sentences. Experimental results reveal that the proposed approach extracts the most relevant sentences when compared to other commercially available text summarizers. Text pre-processing based on word-net and fuzzy analysis is the main part of our work.

1. **Introduction**

The process of automatic text summarization, where a computer automatically creates a summary of a long document, is significantly different from that of human based text summarization since humans can capture and relate deep meaning and themes of text documents. Automation of such a skill is very difficult to implement. Document summarization refers to the task of creating document surrogates that are smaller in size but retain various characteristics of the original document [10].

Nowadays, it is quite common that a keyword-based search on the Internet returns hundreds, or even thousands of hits, by which the user is often confused. Therefore, there is an increasing need for new technologies that can help the user to sift through large volumes of informations and to quickly identify the most relevant documents [6].

Research into automatic text summarization has received considerable attention in the past few years due to the exponential growth in the quantity and complexity of information sources on the internet. Such text summarizers can be used to select the most relevant information from an abundance of text sources that result from a a search engine query [8]. Many summarization models have already been proposed, none of which are entirely based on the document structure. Furthermore, they do not take into account of the fact that the human abstractors extract sentences according to the hierarchical document structure. Related research has shown that human abstractors use ready-made text passages from source documents for summarization. 80% of the sentences in man-made abstracts are closely matched with sentences in source documents. As a result, selection of representative

Received: October 2007; Revised: November 2008 and June 2009; Accepted: March 2010
Key words and phrases: Extraction, Fuzzy logic, Text summarization, Word-net.

sentences is considered a good approximation of summarization. Summarization is mainly the selection of sentences from the source document based on their significance in the document using statistical techniques and techniques based on surface domain-independent linguistic analyses [24]. While abstracts created by professionals involve rewriting of text, automatic summarization of documents has focused on extracting sentences from text so that the overall summary satisfies various criteria including: optimal reduction of text and coverage of document themes [23]. With a large volume of text documents, presenting the user with a summary of each document greatly facilitates the task of finding the desired documents. A compact and concise summary enables the user to quickly get a rough idea of the document's content, and to efficiently identify the documents that are most relevant to his/her needs. Various approaches have been applied to the above-mentioned problems, including statistical learning. Kupiec et al. [17] proposed the first known supervised learning algorithm. Their approach estimated the probability that a sentence should be included in summary, based on its feature values. Chuang and Yang [4] studied several algorithms for extracting segments, such as decision trees and nave Bayes classifiers. These methods perform well for summarizing documents in a specific domain. However, they require a very large amount of training sets to learn accurately. Mani [14] introduced a structured feature. In this method a rhetorical tree structure is built to represent rhetorical relations between sentence segments of the documents for non-structural features. Features in this context are those important ideas which are obtained from the text and can be classified as non-structural features (paragraph location, number of bonus words, number of title words, etc.) and structural features (rhetorical relations between units such as causes, antithesis, conditions, contrast, etc.). Neural networks may present a suitable alternative solution paradigm due to their ability to discover nonlinear mappings as well. [7] The technique proposed in this paper applies human expertise in the form of a set of fuzzy rules and a set of non-structural features to consider the cue feature not only in sentence level but also in paragraph and essay level. Specifically, the parser is designed for selecting sentences based on their attributes and locations in the article using a fuzzy logic inference system. The remarkable ability of fuzzy inference engines in making reasonable decisions in an environment of imprecision and uncertainty makes them particularly suitable for applications that involve risk, uncertainty, ambiguity, and that require flexibility and tolerance to imprecise values. These features make them attractive for automatic text summarization. [25]

2. **Automatic Text Summarization**

Automatic text summarization can be classified into two categories based on their approach: i) summarization based on abstraction, and ii) summarization based on extraction. Most of the work in this area is based on the latter approach. In contrast with abstraction method, which heavily utilizes computation power for natural language processing (NLP) with the inclusion of grammars and lexicons for parsing and generation, this method can be simply viewed as the process of selecting important excerpts (sentences, paragraph, etc.) from the original document

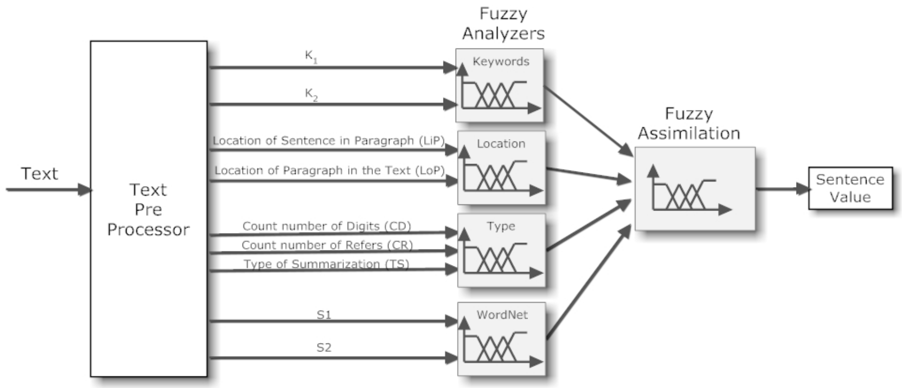

Figure 1. The Overall Structure of the Proposed Text Summarization System

and concatenating them into a more compact form. In other words, extraction is mainly concerned with judging the importance or the indicative power of each sentence in a given document [8]. A query-relevant summary presents the content of the document that is closely related to the initial search query. Creating a query-relevant summary is essentially a process of retrieving the query relevant sentences or passages from the document. When human abstractors extract the sentences, they pay more attention to headings containing some bonus word such as "conclusion", since they consider them as more important to the document therefore more sentences are extracted. The cue feature of a heading sentence is classified as a rhetorical feature [23]. Summarization is not only a function of the input documents but also of the reader's mental state: who the reader is, what his knowledge consists of before reading the summary, and why he wants to know about the input texts. This fact has been long acknowledged by both the psycho-linguistic and the computational-linguistic communities. However, both communities agree that trying to model the reader's mental state is far too complicated, if not entirely impossible. Given this dilemma, most of the computational linguistic research in summarization has assumed that the "reader variable" is a constant and has focused on defining a general notion of salience, valid for all readers [5]. Related research has shown that human abstractors use readymade text passages. 80 % of sentences used in abstracts are closely matched with the source document [cite reference]. As a result, traditional automatic text summarization consists of selection of sentences from source document based on the salient features of the text, such as thematic, location, title, and cue features. Human abstractors extract the topic sentences according to the document structure from top level to low level until they have extracted sufficient information. Traditional automatic summarization techniques adopt the traditional salient features, but they consider the document as a sequence of sentences.

Some of the textual characteristics which have been used in the previous text summarization methods are binary parameters; i.e. they have the value of 0 or 1, which do not function accurately all the time. In our method, we solve this problem by defining the given attributes as fuzzy qualities. This means that each sentence will have the appropriate value between 0 and 1 depending on the particular attribute it has. To optimize the previous text summarization methods using fuzzy logic, we propose several models based on machine learning algorithm. Based on the statistical characteristics of Retrieval Information method, these models rank the sentences of the original text to appear in the summary in the order of importance. In the overall structure of the proposed text summarization system shown in Figure 1, a set of fuzzy analyzers are employed to use information extracted from a given text for evaluating the rank of each sentence of the text regarding the summarization principles. The proposed text summarization system has been implemented in two stages: 1) text pre processing and 2) fuzzy analysis. In the first stage some statistical parameters of the text are computed. In the second stage the fuzzy analyzers compute the rank of the sentences based on the statistical parameters. Finally, the sentences are selected according to the rank and request summarization level. We developed an automatic program based on the proposed approach in MATLAB/Simulink which parses the text into its sentences and identifies the following non-structural features for each sentence:

(1) The number of title words in the sentence,
(2) Whether it is the first sentence in the paragraph,
(3) Whether it is the last sentence in the paragraph,
(4) The number of words in the sentence,
(5) The number of thematic words and keyword synonyms in the sentence.

[8]. The main reasons for using the above features are explained in [2] and [1]. In these papers it is shown that summaries consisting of leading sentences outperform most other methods in this field. It is demonstrated that sentences located at the beginning and the end of the paragraphs are likely to be good summary sentences. Feature 1 indicates the number of title words in a sentence relative to the maximum possible. This is determined by counting the number of matches between the content words in a sentence and the words in the title. Feature 4, length of a sentence, is useful for filtering out short sentences such as datelines and author names commonly found in news articles. Generally, short sentences are unlikely to be included in summaries [9]. This feature is expected to be important because the salience of a sentence according to ISO definition may be affected by the number of words in the sentence also appearing in the title. We aim to use the criteria mentioned above to design fuzzy text analyzers shown in block diagram of the Figure 1. The parameters extracted from the text are given to the fuzzy inference system. The universes of discourse for the input variables are partitioned into several triangular fuzzy sets which have been modified for gaining better performance [22]. The reason for using triangular fuzzy membership function is that the mathematical operations needed for computation of the crisp output applying the fuzzy rules for triangular fuzzy sets are much simpler than other shapes of fuzzy membership functions. Because the

triangular shapes are composed of linear functions with quite simple equations. So, the proposed fuzzy text summarization system does not need huge computational resources and the system summarizes the texts quick enough using the commercial multipurpose computers. This feature makes the proposed system capable to be widely used by users. Performance evaluations on this summarization method are conducted by comparing their summarization outputs with the manual summaries generated by three independent human evaluators. The selection of features plays an important role in determining the type of sentences that will be selected as a part of the summary and therefore, would influence the performance of this fuzzy inference system [17].

3. Fuzzy Analyzers

The remarkable ability of fuzzy inference engines in making reasonable decisions in an environment of imprecision and uncertainty makes them particularly suitable for applications involving risks, uncertainty and ambiguity that require flexibility and tolerance to imprecise values. The main feature of fuzzy logic is that it is able to deal with imprecise linguistic information which makes it attractive for automatic text summarization.

In this paper a fuzzy query-relevant text summarization approach has been proposed to create text summaries by ranking and sentences extraction. Fuzzy analyzers sort all sentences in terms of their importance. They measure sentence relevancies to identify semantically important sentences. This is an attempt to create a summary with a wider coverage of the document's main content and less redundancy. A summary is obtained by choosing a number of top scoring sentences. Fuzzy analyzers evaluate the sentence in various aspects and infer the rank of all sentences of the text. Since there is no clear formula among the mentioned parameters, some if-then rules (called fuzzy rules) have been extracted to describe the relationships among parameters.

The fuzzy rules are formed based on the criteria explained in [2] and [1]. For example any sentence that contains more keywords is more likely to be in the summarized text and gets higher rank. Location of the sentence in the paragraph is important as well. The first sentence of the paragraph gets higher rank. These criteria are organized as four main features which are analyzed with four different fuzzy analyzers: *Keyword* (K), *Location* (L), *Summary-Type* (T) and *Word-net* (W) which are described in sections 3.1 and 3.2.

Fuzzy inference methods are classified in *direct* methods and *indirect* methods. Direct methods, such as *Mamdani* and *Sugeno* are the most commonly used (these two methods are different only in how they obtain the outputs). Indirect methods are more complex. Mamdani fuzzy inference system (FIS) is the most commonly used in applications, due to its simple structure of *min-max* operations. Mamdani type inference expects the output membership functions to be fuzzy sets.

Sugeno method is computationally effective and works well with optimization and adaptive techniques, which makes it very attractive in control problems, particularly for dynamic nonlinear systems.

Mamdani method has been used for aggregation in *Keyword* (K), *Location* (L), *Summary-Type* (T) and *Word-net* (W) fuzzy analyzers. It is used since it is intuitive and well suited for human input while it has widespread acceptance for fuzzy rule designers. Comparing to Sugeno method, which is another commonly used method, the Mamdani is widely accepted for capturing expert knowledge that allows us to describe the expertise in more intuitive, more human-like manner which makes it suitable for text summarization systems. Although Mamdani-type fuzzy inference entails a substantial computational burden, using triangular membership functions the complexity of the computations has been reduced effectively.

After the aggregation process, there is a fuzzy set for each output variable that needs defuzzification. In many cases it is much more efficient to use a single spike as the output membership function rather than a distributed fuzzy set. This is sometimes known as a *singleton* output membership function and it can be thought as a pre-defuzzified fuzzy set. It enhances the efficiency of the defuzzification process, which finds the *centre of area* (Centroid) of a two-dimensional function.

Given the inputs, to compute the output of this FIS six steps have to be followed:

(1) Determining a set of fuzzy rules,
(2) Fuzzifying the inputs using the input membership functions,
(3) Combining the fuzzified inputs according to the fuzzy rules to establish rule strength,
(4) Combining the fuzzified inputs according to the fuzzy rules to establish rule strength,
(5) Finding the consequence of the rule by combining the rule strength and the output membership function,
(6) Defuzzifying the output distribution to get a crisp output[3],[15].

3.1. **Keyword Fuzzy Analyzer.** The *Keyword* fuzzy analyzer computes the effect of the keywords on each sentence. It consists of 34 fuzzy rules derived from non-structural features extracted for each sentence and the perception based knowledge on the parameters which are effective on text summarization (such as the number of keywords in sentence, length of the sentence and number of all keywords). The fuzzy rules of *Keyword* fuzzy analyzer are formed based on the criteria explained in the papers by Brandow [2] and Baxendale [1]. Some samples of these rules are given below [25], [16].

R_{K^1} : *if* (*K1* is Zero) and (*K2* is Zero) then (*Ko* is Zero)
R_{K^2} : *if* (*K1* is Low) and (*K2* is Zero) then (*Ko* is Low)
R_{K^3} : *if* (*K1* is Zero) and (*K2* is Low) then (*Ko* is Low)

where Zero, Low, Medium, High are linguistic values of fuzzy sets for the *K1* and *K2*.

$$K_1 = {}^{n}\!/_{N_k}, K_2 = {}^{n}\!/_{L_s} \tag{1}$$

where *n*, L_s and N_k represent the number of keywords in sentence, length of the sentence and number of all keywords, respectively. The input fuzzy sets of *K1* and *K2* are depicted in Figure 2. The fuzzy set of *Ko*, the output of *Keyword* fuzzy

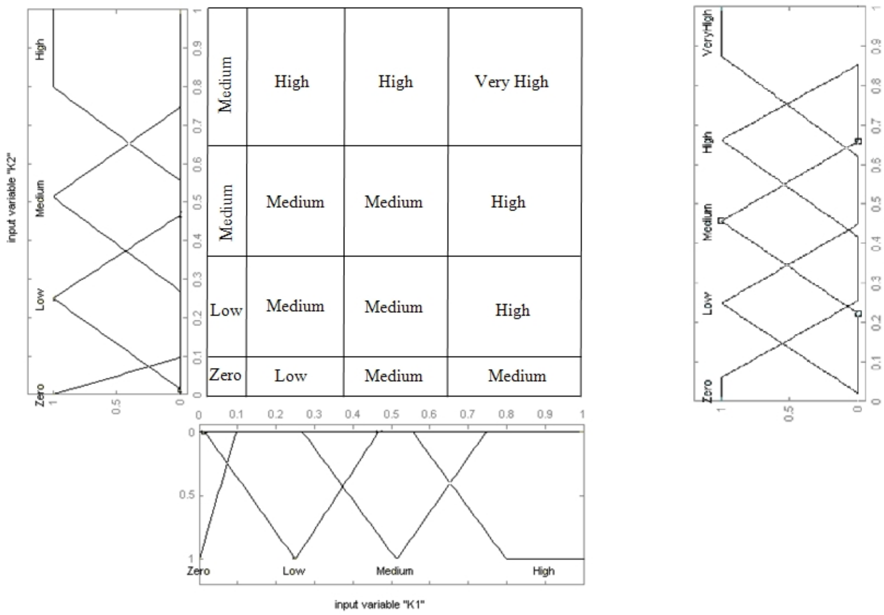

FIGURE 2. Left: The Fuzzy Sets of Keyword Fuzzy Analyzer for K1 and K2 and the Rule Matrix. Right: The Output Fuzzy Sets of Keyword Fuzzy Analyzer

analyzer, is: Zero, Low, Medium, High and Very High. We use a general form to describe these fuzzy rules [21]:

R_{K^i} : *if* (*K1* is X11) and (*K2* is X21) then (*Ko* is Y1), $i = 1 \ldots 34$

where X11 and X21 are triangle-shaped fuzzy numbers [22] and *Y1* is a fuzzy singleton. Based on *Mamdani* implication, membership function is:

$$m^i_{K^i_o} = \mu_{K1^i}(X_{11}).\mu_{K2^i}(X_{21}) \ , \ i = 1 \ldots 34$$

where $\mu_{K1^i}(X_{11})$ and $\mu_{K2^i}(X_{21})$ are membership functions of *K1* and *K2* respectively. The fuzzy sets of *keyword* fuzzy analyzer for *K1* and *K2* and the fuzzy logic rule matrix and the output fuzzy sets of *Keyword* fuzzy analyzer are shown in Figure 2.

A crisp output *Ko* has been obtained using equation (2), via the *center of area* (centroid) method in the defuzzifier:

$$K_o = \frac{\sum_{i=1}^{34} m^i_{K^i_o} \cdot \bar{y}^i_K}{\sum_{i=1}^{34} m^i_{K^i_o}} \tag{2}$$

where $\bar{y}^i_K$ is the centre of the K^i_o area. *Ko* denotes the rank of the sentence got from keyword fuzzy analyzer. The total rank is the summation of all ranks got from the different features. The rest of the features are described briefly in section 3.2.

3.2. **Location, Summary-type and Word-net Fuzzy Analyzers.** In the same way, the general form is used to describe the fuzzy rules of *Location* (L), *Summary-Type* (T) and *Word-net* (W) fuzzy analyzers:

R_{L^i}: (*Location-in-Paragraph* is X12) and (*Location-of-Paragraph* is X22) then (*Lo* is Y2), *i=1...9*

R_{T^i}: (*Number-of-Digits* is X13) and (*Number-of-Refers* is X23) and(*Type-of-Summarization* is X33) then (*To* is Y3), *i=1...27*

R_{W^i}: (*S1* is X14) and (*S2* is X24) then (*Wo* is Y4), *i=1...34*

where *Xij* is triangle-shaped fuzzy number and *Yj* is a fuzzy singleton.

$$S_1 = {}^{n_s}/_{N_k}, S_2 = {}^{n_s}/_{L_s} \tag{3}$$

where n_s, L_s and N_k represent number of keyword synonyms in the sentence, length of sentence and number of all keywords respectively. The linguistic values of the fuzzy sets for the input and output variables are described in Table 1. In this table, *Number-of-Digits* and *Number-of-Refers* are the number of meaningful digits in the sentence and the number of references in the sentence, respectively. The meaningful digits are not the numbering of the paragraphs, figures, tables, etc.

Based on *Mamdani* implication, membership functions are:

$$\begin{gathered} m^i_{L^i} = \mu_{LiP^i}(X_{12}).\mu_{LoP^i}(X_{22}) \text{ , i= 1...9} \\ m^i_{T^i} = \mu_{CD^i}(X_{13}).\mu_{CR^i}(X_{23}).\mu_{TS^i}(X_{33}) \text{ , i= 1...27} \\ m^i_{W^i_o} = \mu_{W^i_1}(X_{11}).\mu_{W^i_2}(X_{21}) \text{ , i= 1...34} \end{gathered}$$

where $\mu_{.^i}(X_{ij})$ is the membership function of parameter *Xij*. The crisp outputs *Lo*, *To* and *Wo* can be computed using equation set (4):

$$L_o = \frac{\sum_{i=1}^{9} m^i_{L_o} \cdot \bar{y}^i_L}{\sum_{i=1}^{9} m^i_{L^i_o}}, T_o = \frac{\sum_{i=1}^{27} m^i_T \cdot \bar{y}^i_T}{\sum_{i=1}^{27} m^i_{T^i}}, W_o = \frac{\sum_{i=1}^{34} m^i_{W^i_o} \cdot \bar{y}^i_W}{\sum_{i=1}^{34} m^i_{W^i_o}} \tag{4}$$

where $\bar{y}^i_L$ and $\bar{y}^i_T$ are the centre of the L^i_o and T^i_o areas respectively.

Fuzzy rules 3-D surfaces are a graphical way to illustrate the fuzzy rules. These graphs show the interpolated extracted rules of each couple of the inputs on X and Y axes. For instance, the 3-D surfaces of fuzzy rules of *Keyword* and *Summary-type* fuzzy analyzers are shown in Figure 3. The values of the signals are normalized.

So far, the rank of each sentence has been computed respecting the features described above. The total rank of the sentence is the summation of all ranks got from each feature for each sentence. The total rank is the measure for the sorting the sentences in terms of importance.

The fuzzy analyzer	**I/O**	**Variable name**	**Fuzzy term sets**
Keyword	input	*K1*	Zero, Low, Medium, High
Keyword	input	*K2*	Zero, Low, Medium, High
Keyword	Output	*Ko*	Zero, Low, Medium, High and Very High
Location	input	*Location-in-Paragraph (LiP)*	First, Middle and End
Location	input	*Location-of-Paragraph (LoP)*	Top, Middle and Down
Location	Output	*Lo*	Zero, Low, Medium, High and Very High
Summary Type	Input	*Number-of-Digits (CD)*	L (Low), Medium (M) and High (H)
Summary Type	Input	*Number-of-Refers (CR)*	L (Low), Medium (M) and High (H)
Summary Type	Input	*Type-of-Summarization (TS)*	Statistical, Normal and Journal
Summary Type	Output	*To*	Zero, Low, Medium, High and Very High
Wordnet	Input	*S1*	Zero, Low, Medium, High
Wordnet	Input	*S2*	Zero, Low, Medium, High
Wordnet	Output	*Wo*	Zero, Low, Medium, High and Very High

Table 1. The Linguistic Values of Fuzzy Sets for the Input and Output Variables

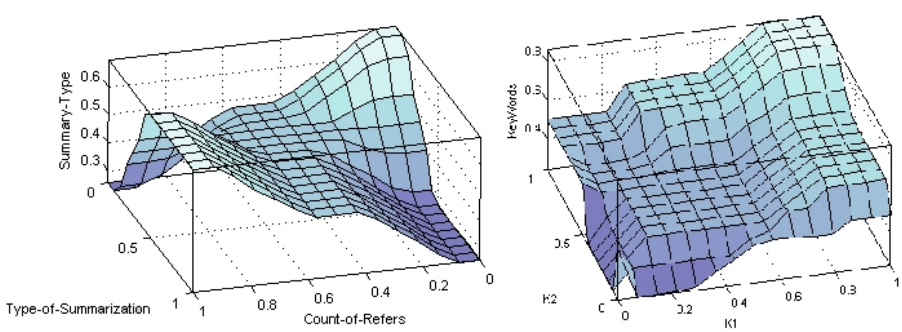

Figure 3. The 3-D Surfaces of Fuzzy Rules of *Keyword* and *Summary-type* Fuzzy Analyzers

4. Simulation and Discussion

For the simulation of the proposed method, a sample text is imported to MATLAB simulation environment. A sample text is shown in the box below:

> As the Internet is growing exponentially, huge amount of information are available online in the websites of libraries and universities. It is difficult to identify the relevant information. The information-overloading problem can be reduced by automatic summarization in 21 century .
> Related research has shown that human abstractors use readymade text passages, 80% sentences in abstract were closely matched with source document or textbook. As a result, traditional automatic text summarization is selection of superior sentences from source document based on the salient features of document, such as thematic, location, title, and cue features.
> Human abstractors extract the topic sentences according to the document structure from top level to low level until they have extracted sufficient information. The traditional automatic summarization techniques adopt the traditional salient features, but they consider the document as a sequence of sentences.
> In the fractal summarization, the traditional salient features are adopted and the hierarchical fractal structure is also considered. The fractal values of the nodes in the hierarchical fractal structure are computed based on the traditional salient features.
> Many summarization models have been proposed previously but None of the models are entirely based on document structure, and they do not take into account of the fact that the human abstractors extract sentences according to the hierarchical document structure.
> In this paper we propose fuzzy query-relevant text summarization method that creates text summaries by ranking and extracting sentences from the original documents and this method uses fuzzy logic to measure sentence relevancies to identify semantically important sentences for summary creations. The method is an attempt to create a summary with a wider coverage of the document's main content and less redundancy.

Keywords extracted from the title of the text (reference sentence) are shown in box below.

information	fuzzy	text	summarization	ranking

For a basic comparison, the sample text is given to some other commercial available text summarizers. All sentences ranked by the proposed method are shown in Table 1(*Rank* column). The sentences ranked by other approaches are shown in Table 1 (*b, c, d* and *e* columns). Column *f* shows the sentences ranked by average of 15 human summarizers.

a) The sentences sorting according to the fuzzy logic based approach proposed in this paper.
b) The sentences sorting using *Microsoft Word Summarizer* tool.
c) The sentences sorting using *Copernic* summarization tool. Using sophisticated statistical and linguistic algorithms, it locates the key concepts and extracts the most relevant sentences, resulting in a Web site or document summary that is a shorter, condensed version of the original text.
d) The sentences sorting using *Pertinence* summarization tool
e) The sentences sorting using *SweSum* automatic text summarizer.
f) The sentences sorting by human (average of 15 different summarizers)

Rank	Sentence	a	b	c	d	e	f
1.3161	'Related research has shown that human abstractors use readymade text passages, 80% sentences in abstract were closely matched with source document or text-book'	1	4	1	2	5	1
1.2864	'As a result, traditional automatic text summarization is selection of superior sentences from source document based on the salient features of document, such as thematic, location, title, and cue features'	2	3	2	1	2	3
1.0866	'In this paper, we propose fuzzy query-relevant text summarization method that creates text summaries by ranking and extracting sentences from the original documents and this method uses fuzzy logic to measure sentence relevancies, to identify semantically important sentences, for summary creations'	3	7	6	4	3	2
1.0344	'The information-overloading problem can be reduced by automatic summarization in century 21'	4	11	8	3	1	4
1.0274	'It is difficult to identify the relevant information'	5	5	9	10	4	7
0.9823	'In the fractal summarization, the traditional salient features are adopted and the hierarchical fractal structure is also considered'	6	6	5	8	7	6
0.9727	'The traditional automatic summarization techniques adopt the traditional salient features, but they consider the document as a sequence of sentences'	7	1	4	5	6	5
0.9633	'Human abstractors extract the topic sentences according to the document structure from top level to low level until they have extracted sufficient information'	8	2	3	6	8	8
0.9432	'Many summarization models have been proposed previously but None of the models are entirely based on document structure, and they do not take into account of the fact that the human abstractors extract sentences according to the hierarchical document structure'	9	8	10	7	9	10
0.7201	'As the Internet is growing exponentially, huge amount of information are available online in the websites of libraries and universities'	10	10	11	9	10	9
0.6733	'The fractal values of the nodes in the hierarchical fractal structure are computed based on the traditional salient features'	11	9	12	11	11	11
0.6733	'The method is an attempt to create a summary with a wider coverage of the document's main content and less redundancy'	12	12	7	12	12	12

Table 2. Sentence Ranking Using the Approach Presented in This Paper and the Other Commercially Text Summarizers

The normalized ranks of 12 sentences using summarization tools mentioned on Table 2 are shown in Figure 4 (a). The blue points represent the rank of the sentences using a fuzzy-logic-based approach. Figure 4 (b) depicts the difference (error) between the ranks obtained from the fuzzy-logic-based approach and the average ranks obtained from the other approaches using the following equation:

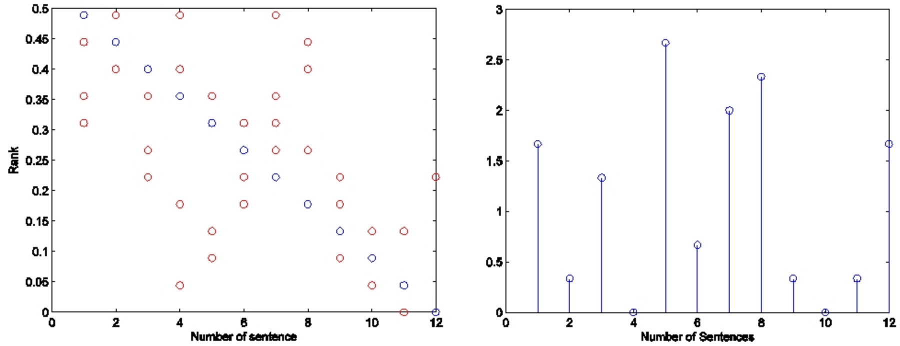

FIGURE 4. (Left) The Normalized Ranks of 12 Sentences, (Right) The Errors of Each Sentence between the Ranks of Fuzzy Logic Based Approach and Average of the Other Approaches

$$E = \sum_{i=1}^{12} \left| R_a(i) - \frac{R_b(i) + R_c(i) + R_d(i) + R_e(i)}{4} \right| \tag{5}$$

where R_a, R_b, R_c, R_d and R_e are the matrices of the ranks for each sentence using the methods *a, b, c, d* and *e* respectively.

The overall average ranking error (difference) between the sentence ranking of proposed method and the sentence ranking average of the other methods is computed as follows:

$$E_{av} = \frac{E}{12} = 1.375 \tag{6}$$

whereE_{av} is the overall average ranking error (difference) between the sentence ranking of proposed method and the sentence ranking average of the other methods. According to equation (6), the ranking error of the fuzzy logic based approach is less than 12% of the sentence ranking average of the other approaches. The proposed fuzzy approach to summarization can be viewed as a category of current systems where some enhancements are applied to make the summarization more similar to the human summarization. There is no statistically significant difference between the proposed fuzzy approach and the other methods in terms of results. The proposed fuzzy approach was evaluated using *rouge* referencing of some human summarizations of the same text as is described in the next section.

5. Rouge Evaluation

Traditionally, evaluation of summarization involves human judgment of different quality metrics including (for example): coherence, conciseness, grammaticality, readability, and content [?]. However, even simple manual evaluation of summaries on a large scale over a few linguistic quality questions and content coverage as in the Document Understanding Conference (DUC) [18] would require over 3,000 hours of human efforts. This is very expensive and difficult to conduct on a frequent basis. Therefore, methods to evaluate summaries automatically have drawn a lot of attention in the summarization research community in recent years. For example, in [20] Saggion et al. proposed three content-based evaluation methods that measure similarity between summaries. These methods include: *cosine similarity*, *unit overlap* (i.e. unigram or bigram), and *longest common subsequence*. However, they do not show how the results of these automatic evaluation methods correlate to human judgments. Following the successful application of automatic evaluation methods such as BLEU [19] in machine translation evaluation, Lin and Hovy in [13] showed that methods similar to BLEU, i.e. n-gram co-occurrence statistics, could be applied to evaluate summaries.

5.1. **Rouge-L Evaluation.** In this section an automatic evaluation of summaries is used. A sequence $Z = [z1,\ z2,\ \ldots,\ zn]$ is a subsequence of another sequence $X = [x1,\ x2,\ \ldots,\ xm]$, if there exists a strict increasing sequence $[i1,\ i2,\ \ldots,\ ik]$ of indices of X such that for all $j = 1,\ 2,\ \ldots,\ k$, we have $xij = zj$. Given two sequences X and Y, the longest common subsequence (LCS) of X and Y is a common subsequence with maximum length.

To apply LCS in summarization evaluation, a summary sentence is viewed as a sequence of words. The intuition is that the longer the LCS of two summary sentences is, the more similar the two summaries are. When applying to summary-level, we take the union LCS matches between a reference summary sentence, *ri* and every candidate summary sentence, *cj*. Given a reference summary of u sentences containing a total of m words and a candidate summary of v sentences containing a total of n words, the summary-level LCS-based F-measure can be computed as follows:

$$P_{lcs} = \frac{\sum_{i=1}^{u} LCS_{\bigcup}(r_i, C)}{n} \tag{7}$$

$$R_{lcs} = \frac{\sum_{i=1}^{u} LCS_{\bigcup}(r_i, C)}{m} \tag{8}$$

$$F_{lcs} = \frac{(1+\beta^2)R_{lcs}P_{lcs}}{R_{lcs} + \beta^2 P_{lcs}} \tag{9}$$

where β could be set to a very big number ($\rightarrow \infty$) in DUC, i.e. only R_{lcs} is considered or it can be set to 1. $LCS_{\bigcup}(r_i, C)$ is the LCS score of the union longest common subsequence between reference sentence r_i and candidate summary C. The LCS-based F-measure, i.e. Equation (5.1) is called ROUGE-L [11]. Notice that ROUGE-L is 1 when $r_i = C$; while ROUGE-L is zero when $LCS\ (r_i,\ C) = 0$.

A good automatic metric should have high *Rlcs (recall)* and *Plcs (precision)*. This implies that if a statistical test indicates a significant difference between two runs using the automatic metric, then there is probably also a significant difference in the manual evaluation. This would be very useful to gauge during the system development cycle if an improvement is really significant or not [8], [12].

In an experimental comparison with referencing to the average of 15 human summarizers shown in Table 2 column (f), the accuracy of this method was found to be 84% (F_{lcs}=*0.84*) of that of a human summarizer. The speed of the summarizer was also highly satisfactory for web applications due to the use of fuzzy inference which was programmed in parallel codes.

5.2. **ROUGE-N: N-gram Co-occurrence Statistics.** Formally, ROUGE-N is an n-gram recall between a candidate summary and a set of reference summaries. ROUGE-N is computed as follows:

$$ROUGE\_N = \frac{\sum_{S \in \{Reference-Summaries\}} \sum_{gram_n \in S} Count_{match}(gram_n)}{\sum_{S \in \{Reference-Summaries\}} \sum_{gram_n \in S} Count(gram_n)} \quad (10)$$

where n stands for the length of the n-gram, $gram_n$, and $Count_{match}$ $(gram_n)$ is the maximum number of n-grams co-occurring in a candidate summary and a set of reference summaries. It is clear that ROUGE-N is a recall-related measure because the denominator of the equation (10) is the total sum of the number of n-grams occurring at the reference summary side. A closely related measure, BLEU, used in automatic evaluation of machine translation, is a precision-based measure. BLEU measures how well a candidate translation matches a set of reference translations by counting the percentage of n-grams in the candidate translation overlapping with the references.

Note that the number of n-grams in the denominator of the ROUGE-N formula increases as we add more references. This is intuitive and reasonable because there might exist multiple good summaries. Every time we add a reference into the pool, we expand the space of alternative summaries. By controlling what types of references we add to the reference pool, we can design evaluations that focus on different aspects of summarization. Also note that the numerator sums over all reference summaries.

5.3. **The Results of All Models Trained on English Data and Tested on Duc 2003 Data.** In this experiment, we train all previously mentioned models on the 10 English features (using the same 50 English articles) and test these models on the DUC 2003 data to investigate the proposed system performance on a newswire data. We have created new extractive reference summaries of the DUC 2003 testing data by measuring the similarity (vocabulary overlap) between each sentence and the associated reference single document summary. Then we rank each document sentences based on this similarity value. A set of sentences is specified as a reference summary for each document based on the compression ratio. Table 3 shows the results of all models for the DUC 2003 testing data based on the average Rouge-1 score.

6. Discussion

As shown in Table 4, the precision decreased slightly when the models were trained on one language and tested on the other language. However, the decrease in precision was not significant. Therefore, it is possible to train some models with some features and use them for another language. Moreover, it is evident from Table 5 that this approach can be extended to the genre of newswire text.

Compression rate (CR)	**10%**	**95% Confidence interval**	**20%**	**95% Confidence interval**	**30%**	**95% Confidence interval**
Av_Rouge-1(*Copernic* summarization tool)	0.4326	0.4142, 0.4509	0.4435	0.4251, 0.4618	0.4534	0.4350, 0.4717
Av_Rouge-1(*Microsoft Word Summarizer* tool)	0.4302	0.4106, 0.4497	0.4314	0.4118, 0.4509	0.4357	0.4161, 0.4552
Av_Rouge-1(*Pertinence* summarization tool)	0.4543	0.4309, 0.4777	0.4626	0.4392, 0.4860	0.4678	0.4444, 0.4912
Av_Rouge-1(*SweSum* automatic text summarizer)	0.4587	0.4330, 0.4844	0.4756	0.4499, 0.5013	0.4793	0.4536, 0.5050
Av_Rouge-1(Fuzzy Model)	0.6075	0.5778, 0.6372	0.6124	0.5827, 0.6421	0.6257	0.5960, 0.6554

Table 3. All Models Performance Evaluation Based on the Average Rouge-1 Score (Duc 2003 Testing Data)

Compression rate (CR)	**10%**	**95% Confidence interval**	**20%**	**95% Confidence interval**	**30%**	**95% Confidence interval**
P(*Copernic* summarization tool)	0.4243	0.4070, 0.4417	0.4321	0.4148, 0.4495	0.4376	0.4203, 0.4550
P(*Microsoft Word Summarizer* tool)	0.4153	0.3915, 0.4392	0.4038	0.3800, 0.4277	0.4063	0.3825, 0.4302
P(*Pertinence* summarization tool)	0.4464	0.4222, 0.4706	0.4412	0.4170, 0.4654	0.4453	0.4211, 0.4695
P(*SweSum* automatic text summarizer)	0.4453	0.4274, 0.4632	0.4586	0.4407, 0.4765	0.4603	0.4424, 0.4782
P(Fuzzy Model)	0.5923	0.5750, 0.6097	0.5976	0.5803, 0.615	0.6092	0.5919, 0.6266

Table 4. All Models Performance Evaluation Based on Precision (English Testing Data)

Compression rate (CR)	**10%**	**95% Confidence interval**	**20%**	**95% Confidence interval**	**30%**	**95% Confidence interval**
P(*Copernic* summarization tool)	0.4152	0.3968, 0.4335	0.4236	0.4053, 0.4419	0.4335	0.4132, 0.4538
P(*Microsoft Word Summarizer* tool)	0.4098	0.3869, 0.4326	0.4021	0.3793, 0.4249	0.4021	0.3803, 0.4239
P(*Pertinence* summarization tool)	0.4383	0.4171, 0.4595	0.4403	0.4191, 0.4615	0.4423	0.4191, 0.4655
P(*SweSum* automatic text summarizer)	0.4438	0.4249, 0.4627	0.4526	0.4337, 0.4715	0.4543	0.4344, 0.4742
P(Fuzzy Model)	0.5902	0.5718, 0.6085	0.5936	0.5753, 0.6119	0.6046	0.5873, 0.6219

Table 5. All Models Performance Evaluation Based on Precision (Duc 2003 Testing Data)

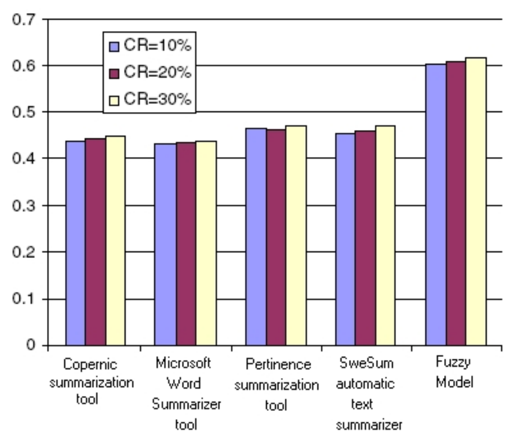

Figure 5. The Results Associated with All Models for Different Cr (English Case)

Figure 5 shows the total system performance in terms of precision for different compression rates in case of all models for English articles. It is clear from the figures that *Fuzzy Model* has a good capability to model arbitrary densities and therefore yields the best results. The *SweSum* automatic text summarizer is more precise than the Pertinence summarization tool, which is better than the Microsoft Word in precision. And finally, the Microsoft Word Summarizer tool is better than the *Copernic* summarization tool. It is also clear that decreasing the training data size to half (English case) does not severely affect the total system performance.

7. Conclusion

In this paper, we present a new approach for creating summaries using fuzzy inference system. The analysis of the parameters which are important in summarization was done by a number of fuzzy-logic-based analyzers. This text summarization system consists of 1) text pre-processor which extracts different information needed for fuzzy analysis from the text using word-net database and 2) an analyzers which contain fuzzy-logic-based inference systems to compute weighted score of each sentence in the text. The scores of relevance have been ranked. Starting with the highest score, the analyzer includes in the summary the sentences for which the relevance score is higher than the threshold value set. The process continues until the ratio of compression satisfies the limitation set initially. A simulation model was developed in MATLAB for this approach. The word-net MATLAB interface was used to extract further information from the original text. The advantage of this method is that linguistic variables and human perception are taken into consideration. Statistical comparison to current commercially-available summarizers shows that the proposed fuzzy approach can be viewed as a category of the current summarizer systems where using fuzzy rules some enhancements are applied to make it more similar to the human summarization. Evaluation using Rouge indicates the advantage of this approach in comparison to referencing human summarizations. The weakness of the proposed fuzzy summarizer is that the process of designing fuzzy rules, which have to cover all the relationships among the parameters, is quite time consuming.

References

[1] P. B. Baxendale, *Machine made index for technical literature: an experiment*, IBM Journal of Research and Development, **2(4)** (1958), 354-361.

[2] R. Brandow, K. Mitlze and L. Rau, *Automatic condensation of electronic Publication by sentence election*, Information Processing and Management, **31(5)** (1995), 675-685.

[3] J. J. Buckley, K. D. Reilly and L. J. Jowers, *Simulating continuous fuzzy systems: I*, Iranian Journal of Fuzzy Systems, **2(1)** (2005), 1-18.

[4] W. T. Chuang and J. Yang, *Extracting sentences segments for text summarization: a machine learning approaches*, Proceedings of the 23th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Athens, Greece, (2000), 125-159.

[5] N. Elhadad, *User-sensitive text summarization thesis summary*, Thesis Summary, American Association for Artificial Intelligence, USA, 2004.

[6] Y. Gong and X. Liu, *Creating generic text summaries*, IEEE, 0-7695-1263-1/01, (2001), 391-407.

[7] K. Kaikhah, *Automatic text summarization with NNs*, Second IEEE International Conference on Intelligent Systems, June (2004), 40-44.

[8] A. Kiani-B, M. R. Akbarzadeh-T and M. H. Moeinzadeh, *Intelligent extractive text summarization using fuzzy inference systems*, 1-4244-0457-6/06, IEEE, (2001), 1-4.

[9] J. Kupiec, J. Pederson and F. Chen, *A trainable document summarizer*, Proceedings of the 18th Annual international ACM SIGIR Confluence on Research and Development in Information Retrieval, Seattle, Washington, (1995), 68-73.

[10] J. Leskovec, M. Grobelnik and N. Milic-Frayling, *Learning semantic graph mapping for document summarization*, Proceedings of ECML/PKDD-2004 Workshop on Knowledge Discovery and Ontologies, KDO-2004, Pisa, Italy.

[11] C. Y. Lin, *ROUGE: a package for automatic evaluation of summaries*, Proceedings of Workshop on Text Summarization Branches Out, Post-conference Workshop of ACL, Spain, 2004.

[12] C. Y. Lin and E. Hovy, *Automatic evaluation of summaries using n-gram co-occurrence statistics*, Proceedings of the Human Technology conference (HLT-NAACL-2003), Canada, (2003), 71-78.

[13] C. Y. Lin and E. H. Hovy, *Automatic evaluation of summaries using n-gram co-occurrence statistics*, Proceedings of Language Technology Conference (HLT-NAACL 2003), Edmonton, Canada, (2003), 287-292.

[14] I. Mani, *Advances in automatic summarization*, John Benjamins Publishing Company, (2001), 129-165.

[15] E. G. Mansoori, M. J. Zolghadri and S. D. Katebi, *Using distribution of data to enhance performance of fuzzy classification systems*, Iranian Journal of Fuzzy Systems, **4(1)** (2007), 22-36.

[16] G. A. Miller, R. Beckwith, C. Fellbaum, D. Gross, and K. Miller, *Five papers on wordnet*, Technical Report, Princeton University, (1993), 3-12.

[17] T. Nomoto and Y. Matsumoto, *A new approach to unsupervised text summarization*, SIGIR, ACM, New Orleans, Louisiana, USA, (2001), 26-34.

[18] P. Over and J. Yen, *An introduction to duc 2003 - intrinsic evaluation of generic news text summarization systems*, http:// wwwnlpir.nist.gov/ projects/ duc/ pubs/ 2003slides/ duc2003intro.pdf, 2003.

[19] K. Papineni, S. Roukos, T. Ward and W. J. Zhu, *BLEU: A method for automatic evaluation of machine translation*, IBM Research Report RC22176 (W0109-022), 2001.

[20] H. Saggion, D. Radev, S. Teufel and W. Lam, *Meta-evaluation of summaries in a cross-lingual environment using content-based metrics*, Proceedings of COLING, Taipei, Taiwan, 2002.

[21] A. K. Shaymal and M. Pal, *Triangular fuzzy matrices*, Iranian Journal of Fuzzy Systems, **4(1)** (2007), 75-87.

[22] L. X. Wang, *A cource in fuzzy system and control*, Prentice Hall, Englewood Cliffs, Nj. ISBN-13: 978-01354088271998.

[23] C. C. Yang and F. L. Wang, *Fractal summarization: summarization based on fractal theory*, SIGIR, ACM 1-58113-646, Toronto, CA, (2003), 391-392.

[24] C. C. Yang and F. L. Wang, *Hierarchical summarization of large documents*, Journal of the American Society for Information Science and Technology, **59(6)** (2008), 887-902.

[25] L. A. Zadeh, *Fuzzy sets as a basis for a theory of possibility*, Fuzzy Sets and Systems, Elsevier, Holland, (1999), 9-34.

Farshad Kyoomarsi, Islamic Azad University of Shahrekord branch, Shahrekord, Iran

Hamid Khosravi, Shahid Bahonar University of Kerman, International Center for Science and High Technology and Environmental Sciences, Kerman, Iran

Esfandiar Eslami, Shahid Bahonar University of Kerman, The centre of Excellence for Fuzzy system and applications, Kerman, Iran

Mohsen Davoudi*, Department of Energy, Electrical Engineering division, Politecnico di Milano, Milan, Italy

*Corresponding author