DOI: 10.24996/ijs.2017.58.3A.14

ISSN: 0067-2904

TOPSIS with Multiple Linear Regression for Multi-Document Text Summarization

Suhad Malallah[1], Zuhair Hussein Ali[*2]
[1]Computer Science Department, University of Technology, Baghdad, Iraq.
[2] Computer Science Department, College of Education, Al- Mustansiriya University, Baghdad, Iraq.

Abstract

The huge amount of information in the internet makes rapid need of text summarization. Text summarization is the process of selecting important sentences from documents with keeping the main idea of the original documents. This paper proposes a method depends on Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS). The first step in our model is based on extracting seven features for each sentence in the documents set. Multiple Linear Regression (MLR) is then used to assign a weight for the selected features. Then TOPSIS method applied to rank the sentences. The sentences with high scores will be selected to be included in the generated summary. The proposed model is evaluated using dataset supplied by the Text Analysis Conference (TAC-2011) for English documents. The performance of the proposed model is evaluated using Recall-Oriented Understudy for Gisting Evaluation (ROUGE) metric. The obtained results support the effectiveness of the proposed model.

Keywords: weight feature, Muliple Linear Regression, TOPSIS,PIS,NIS.

Topsis مع الانحدار الخطي المتعدد لتلخيص النصوص المتعدده

سهاد مال الله[1] ، زهير حسين علي[*2]
[1]قسم علوم الحاسبات، الجامعه التكنلوجيه، بغداد، العراق.
[2]قسم علوم الحاسبات، كلية التربيه، الجامعه المستنصريه، بغداد، العراق.

الخلاصه

بالنظر للكميات الكبيرة الموجودة من المعلومات في الانترنت ادى الى الحاجة الضرورية لتلخيص المعلومات. أن عملية تلخيص المعلومات تتضمن أستخراج الجمل المهمه من النصوص مع المحافظة على الافكار الرئيسية للنصوص الملخصه. هذا البحث يقترح طريقة تعتمد على تقنية ترتيب الافضلية عن طريق التشابه الى الحل المثالي(TOPSIS). الخطوة الاولى في موديلنا المقترح تعتمد على أستخراج سبعة خصائص لكل جملة من جمل النصوص المراد تلخيصها. بعدها تم طريقة أستخدام الانحدار الخطي المتعدد لغرض تعيين أوزان للخصائص المختارة.

ثم تطبق طريقة TOPSISلغرض ترتيب الجمل. يتم أختيار الجمل ذات الدرجة الاعلى لغرض تضمينها ضمن الملخص المتكون.تم أستخدام قاعدة بيانات (TAC–2011) للغة الانكليزية. أختبرت النتائج باستخدام برنامج ROUGE أثبتت النتائج كفاءه النظام المقترح.

*Email: zuhair72h@yahoo.com

1. Introduction

According to the fast development of information-communication technologies, enormous quantity of documents have been created and put together in the World Wide Web. The huge amount of documents makes it difficult for the user to get useful information [1]. To deal with such problem of information overload, Automatic Text Summarization (ATS) has been used as a solution. ATS is the process of generating a single document summary from a set of documents or from a single document without losing its main ideas [2]. This process helps users to the general review of all related documents and interested issues with understanding the main content of the summarized documents; this process also helps to reduce the time needed to get these briefs. Rely on the amount of document to be summarized ATS can be classified as a Single Document summarization (SDS) or Multi Document summarization (MDS). In SDS only one document can be summarized into shorter one, whereas in MDS a set of related documents with same topic is summarized into one shorter summary [3]. Summarization methods, also, can be classified as abstractive summarization and extractive summarization. Aabstractive summarization depends on Natural Language Processing (NLP) strategies, which request deep understanding of NLP techniques to analyze the documents sentences and paragraphs, since some changes have to be done to the selected sentences. Whereas in the extractive summarization, no change is applied to the sentences which are selected to be included in the final summary[4]. Thus abstractive summarization seems to be more difficult and time-consuming than extractive summarization [5]. Also summarization can be categorized as query summarization and generic summarization. In the query based summarization a summary was generated according to the user query, where the documents searched to match with the user query [6]. While generic summarization creates a summary which include the main content of the documents. One of the most challenges for the generic summarization is that no topic or query available for the summarization process [7].

2. Related Works

ATS reduces a large number of text documents to a smaller set of sentences which explain the main ideas of these documents. Specialists in NLP are more interested to discover new methods for summarizing and exploring a variety of models to come up with perfect summarization. In this section we investigate some of these methods [8].

In [9] the authors suggested a method for calculating the weights of the selected features. Five different features were used, the first two are structural features in which consist of more than simple features, while the remaining three features are simple features. These five selected features are used as input parameters to the particle swarm optimization (PSO) used to train these features and assign a weight to each one of them. Their results showed that structural features got average weight higher than simple features. In [10] the authors suggested a method based on selecting five features. These features are: sentence position, sentence length, numerical information, thematic words and title feature. The pseudo genetic algorithm was used to train the dataset and assign a weight to each feature. Their results showed that the importance of these features are in the following order title feature, sentence position, thematic words, sentence length and numerical information. In [11], a set of features were extracted for each setence ; this set was used as input to a model consist of three functions: Cellular Learning Automata (CLA), PSO, and fuzzy logic. The CLA was used to calculate the similarity between sentences to reduce the redundancy. While the PSO was used to set a weight for each feature, then the fuzzy logic was used to give scores to the sentences, these scored sentences were arranged in descending order, and the sentence with higher score was selected to be included in the created summary. In [12], the authors proposed a method based on formulating the problem of MDS as a multi-objective optimization (MOO). Two main objective functions were formulated these are redundancy reduction and content coverage. The redundancy reduction was computed using cosine similarity between each sentence in the dataset, whereas the content coverage was computed using the cosine similarity between each sentence with the mean of document collection. Evolutionary Algorithm was used to combine these two objective functions with the aim to minimize the first objective and maximize the second objective function. Good results are obtained from their method.

The fundamental objective of document summarization is the extraction of suitable and pertinent sentences from the input document(s). A technique to acquire the significant sentences is through assigning a weight for each sentence which indicates the salience of a sentence for selection to the summary and then selecting the top ones [13]. In this paper a method for extracting generic MDS for

English text is proposed which depends on extracting seven features for each sentence in the documents, then a mathematical model is used for assigning a weight for each feature. The mathematical model is based on Multiple Linear Regression (MLR). The weights of the selected features are used as input to the TOPSIS algorithm. The TOPSIS uses both: the selected features and their calculated weights to rank the sentences. We have utilized Text Analysis Conference (TAC-2011) dataset to assess the summarized results.

3. Problem Statement and Formulation

To produce a good summary for any MDS system two issues must be considered. These issues are

1-Relevancey: can be defined as the goodness of information included in the created summary. A summary considered as relevant if it includes many information relevant to the main topic of the documents.

2-Redundancy: The generated summary should include less redundant information to cover most of the relevant topics.

Formally, given a corpus which consists of many clusters, each cluster contains a set of documents called D with the same topic. The set D can be defined as D= $\{d_1, d_2,\dots, d_n\}$ where n is the number of distinct document in D. Each D can be represented by a set of sentences called S_i, i.e D= $\{S_i \mid 1<=i<=M\}$where M represents the total number of sentences in the set D.

Our goal is to find a subset of set D called A i.e. $A \subset D$ that satisfies both objectives: relevancy maximization and redundancy reduction.

4. Basic Concepts

There are two main stages: Preprocessing and feature extraction.

4.1 Preprocessing

There are four steps in this stage.

A- **Sentence segmentation**: which can be done by splitting sentences according to the dot between them.

B- **Tokenization**: Is the process of splitting sentence into words

C- **Stop Words Removal**: Words which don't give the necessary information for identifying significant meaning of the document content and appear frequently are removed. There are a variety of methods used for specifying such stop words list. . Presently, a number of English stop word list is usually used to help text summarization process

D- **Stemming**: is the process of producing root of the word, in This paper word stemming is performed using Porter's stemming algorithm [14].

4.2 Features Extraction

An essential part of ATS is computing features score for every sentence. The features include: sentence position, sentence length, numerical data, thematic word, title word, proper noun and centroid value [15].

A- Sentence Position (SP): higher score will be given to the first sentence; the score decreases according to the sentence position in the document. This feature can be computed according to Eq. (1).

$$F1(s_i) = \frac{N - i + 1}{N} \tag{1}$$

Where i is the position of the sentence (s) in a document of N sentences

B- Sentence length (SL): This feature is computed by dividing the sentence length by the length of the longest sentence in the document as in Eq. (2).

$$F2(s_i) = \frac{L(s_i)}{L_{\max}} \tag{2}$$

Where $L(s_i)$ is the length of sentence s_i and L_{max} is the length of longest sentence in the document.

C- Numerical data (ND): has important information to be included in the summary. This feature is calculated by dividing the number of numerical data in the sentence by the sentence length as in Eq. (3).

$$F3(s_i) = \frac{Num(s_i)}{L(s_i)} \tag{3}$$

Where $Num(s_i)$ is the number of numerical data in the sentence (s_i)
D- Thematic Words (TW): terms that appear most frequently in the document. This feature can be calculated by computing the repetition of all terms in the document, then only (K) terms with the highest repetition is selected, in this work, This feature is calculated by dividing the number of thematic words in the sentence by the maximum number of thematic words in the document as expressed in Eq. (4).

$$F4(s_i) = \frac{TW(s_i)}{TW_{\max}} \quad (4)$$

TW is the number of thematic words in the sentence.
TW_{max} maximum thematic words in the sentences.
E-Title Feature (TF): This feature is important when summarizing the document. The score is calculated as in Eq. (5).

$$F5(S_i) = \frac{No.of\ TF}{L(Title)} \quad (5)$$

Where TF is the words that exists in both: S_i and Title
F-Proper Noun (PN): The sentence is important if it includes the maximum number of proper nouns [16]. This feature is calculated as in Eq. (6)

$$F6(s_i) = \frac{PN(s_i)}{L(S_i)} \quad (6)$$

Where PN is the number of proper nouns in a sentence s_i.
G-Centroid value (CV): Is a feature used to specify salient sentences in the multiple documents [17]. This feature can be calculated as follows

$$F7(S_i) = \sum_{i=1}^{L(S_i)} C_w \quad (7)$$

$$C_w = TF * IDF \quad (8)$$

$$IDF = \log\left[\frac{n}{nw}\right] \quad (9)$$

Where n total number of documents.
nw number of documents containing the given word.
F7 is the centroid value of the sentence S_i
Cw is the centroid value of the word.
TF is the term frequency which represents the frequency of a given term in the document.
IDF is the inverse term frequency computed by division of the total number of documents over the number of documents including the given term.

The division excluded from all the seven features to produce non normalized features that used directly in the TOPSIS algorithms.

5. The proposed method

There are three main stages in the proposed method. The first stage includes how to compute the weights for the extracted features. In the second stage the computed weights of stage one are used as input to TOPSIS algorithm. The TOPSIS algorithm is used to rank all sentences, then higher score sentences will be chosen to be included in the final summary. Third stage includes removing redundancy. Figure-1 shows the block diagram of the proposed method.

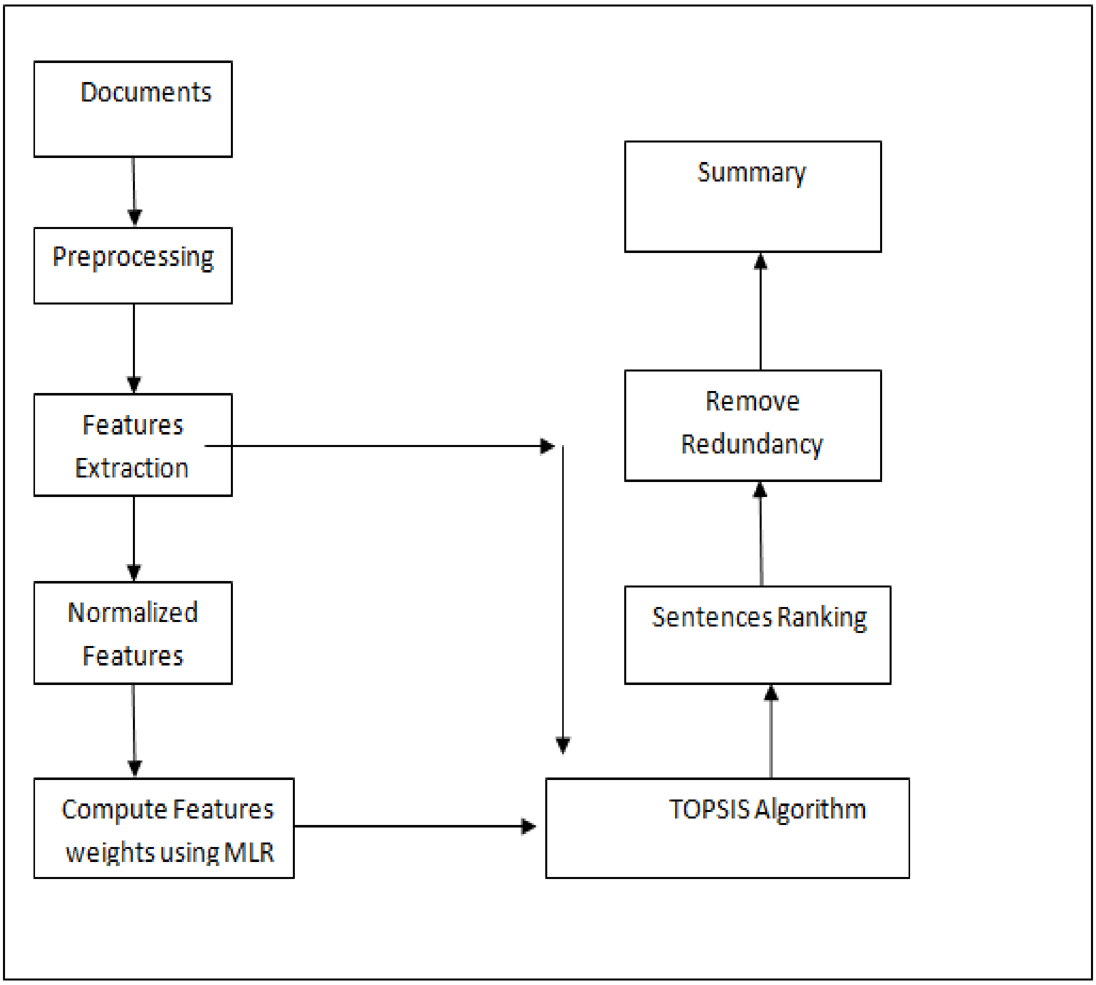

Figure1- Block Diagram for Proposed MDS.

5.1 Multiple Linear Regression

MLR is a statistical method for formulating the relationship between the independent variables and a dependent variable, where there are two or more independent variables, but only one dependent variable [18]. MLR can be formulated as in Eq (10)

$$Y= W_0+W_1 X_1+W_2 X_2+\text{---}+X_m W_m \qquad (10)$$

Where

[Y] is the output vector (dependent variable).

[W] feature weight vector.

[X] The extracted features (independent variables).

The regression model can be represented in a matrix form as follows.

$$\begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ \\ . \\ Y_p \end{bmatrix} = \begin{bmatrix} X_{1,1} X_{1,2} X_{1,3} X_{1,4} X_{1,5} X_{1,6} X_{1,7} \\ X_{2,1} X_{2,2} X_{2,3} X_{2,4} X_{2,5} X_{2,6} X_{2,7} \\ .. \\ X_{p,1} X_{p,2} X_{p,3} X_{p,4} X_{p,5} X_{p,6} X_{p,7} \end{bmatrix} = \begin{bmatrix} W_1 \\ W_2 \\ W_3 \\ W_4 \\ W_5 \\ W_6 \\ W_7 \end{bmatrix}$$

Where p is the number of sentences from the collected document data set. To estimate the weights for the extracted features we must train our model. There are 70 documents from the TAC-2011 dataset used for the training mode. The seven extracted features (X_1, X_2,…, X_7) that were described in section 5 are used as input to the model. The desired output Y can be computed using cosine similarity between all sentences from selected trained documents and each sentence from the manually summarized documents. As in Eq (11)

$$Similarity\,(A,B) = \frac{\sum_{i=1}^{z} A_i B_i}{\sqrt{\sum_{i=1}^{z}(A_i)^2} * \sqrt{\sum_{i=1}^{z}(B_i)^2}} \tag{11}$$

Where A sentence from trained documents
B sentence from manually summarized documents.
Each sentence (A) will be compared with all sentences (B) and higher score value assign to Y_i. The score of the equation (11) range from (0) to (1), zero score means there is no matching between A and B whereas one means A identical to B. Thus we have values of (X_i) and (Y_i) for equation (10).

Our goal is to estimate the values of (W_i) which represent the weights of the selected features [19]. W will be calculated using Eq.(12) .

$$W=(X.X^t)^{-1}.X^tY \tag{12}$$

5.2 TOPSIS Method

TOPSIS is a decision making method that developed by Hwang and Yoon in 1981[20], in 1993. It was more developed by Hwang, Lai and Liu[21]. TOPSIS is an efficient method for Multi-Attribute Decision Making, that is used for ranking, evaluating and selecting the most suitable alternative among various alternatives. TOPSIS depends on choosing an alternative that has nearest distance to the Positive Ideal Solution (PIS) and farthest distance from Negative Ideal Solution [22]. It can help in choosing features that assist to decide which alternatives are the most appropriate for the problems. The weights and scores are the two essential evaluation parameters used to make a decision [23].
In our proposed method the scores and weights of the features are computed mathematically (as described in section 5 and section 6.1) to be used in ranking the sentences. This makes TOPSIS suitable for MDS where there are many criteria (features) and we have to choose the most suitable decision.

5.3 TOPSIS as Summary Generation

In this paper TOPSIS will be employed in MDS to propose a mathematical model for ranking sentences and choose the most suitable ones. To create a decision matrix for TOPSIS seven features and M sentences are used as shown in Table- 1.

Table 1 - Decision Matrix for TOPSIS method

	F1	F2	F3	F4	F5	F6	F7
S_1	$X_{1,1}$	$X_{1,2}$	$X_{1,3}$	.	.	.	$X_{1,7}$
S_2	$X_{2,1}$	$X_{2,2}$	$X_{2,3}$	.	.	$X_{2,6}$	$X_{2,7}$
.	.	.	.	.	.	.	.
S_M	$X_{M,1}$	$X_{M,2}$	$X_{M,3}$	.	.	.	$X_{M,7}$

$X_{i,j}$ is the feature value where, i=1,..,M and j=1,..,7.
Our goal is to get the best sentences that are nearest to the PIS and far from NIS. Each attribute in the decision matrix is arranged either in increasing order or decreasing order. In the proposed method, the attributes are arranged only in decreasing order from the highest to lowest one to get sentences with the highest score [24]. Algorithm (1) describes the main steps of TOPSIS technique

TOPSIS Algorithm

Step1 :input decision matrix {section 5.3}

Output sentences in descending order

Step2: Normalized a decision matrix by $R_{i,j} = \dfrac{x_{i,j}}{\sqrt{\sum_{i=1}^{n}(x_{i,j})^2}}$

Step3: Use MLR to construct the feature weights vector W_j

Step 4: Construct the Weighted Normalized Decision Matrix by multiplying each column by W_j to get $V_{i,j}$

Step5: Determine the ideal solution for each column A^*{The highest value in the column V_j^+}

Step6: Determine the Negative solution for each column A^-{The lowest value in the column V_j^-}

Step7: Determined the PIS

$$S_i^{+} = \sqrt{\sum_{j=1}^{m}(V_{i,j} - V_j^{+})^2}$$

Step8:Determine the NIS $S_i - = \sqrt{\sum_{j=1}^{m}(V_{i,j} - V_j^{-})^2}$

Step9 :calculate closness to ideal solution $C_i = S^{-}{}_i / (S^{+}{}_i + S^{-}{}_i)$

Step10:rank all sentences according to the results of step9

By **algorithm 1-** all the sentences arranged in descending order depending on their score.

5.4 Remove Redundancy

This stage is very necessary for MDS. There are many documents with the same topic, some sentences may be repeated in more than one document. A technique is required to remove the redundant sentences from the generated summary, which allows the final summary to include the most important ideas for the summarized documents [15]. The cosine similarity as explained in Eq. (11) is used to compute the similarity between two sentences and exclude the sentence from a final summary when the similarity between them is more than a specified threshold. The following algorithm (2) illustrates reducing redundancy and generating summary in the proposed MDS model.

Summary generating algorithm

input 1- set of ranked sentences in descending order from topsis algorithm called scored_sent

2- Max summary size called Max_size

output generated summary called summary
Step1: let summary =[]
Size=0
No_of_sen=0
Step2 : from Scored_sent select Si with highest score
Step3 :Flag=false
for j from 1 to No_of_sen
compare Si with Sj{Sj sentence from summary} according to Eq.(11)
If (Similarity(Si,Sj) >threshold) then flag=true
Step4:if (flag) delete Si go to step2
Else Put Si in the summary
Size=size+count_words (Si)
NO_of_sen=No_of_sen+1
Step5: if size<max_size goto step2
Else end

6. Dataset and Evaluation Metrics

The dataset used in our experiment is TAC-2011 which consists of a document set written in seven languages (English, Arabic, Greek, Czech, French, Hindi, Hebrew). For each language (10) topics are used each of (10) documents. Summarization of (10) pre evaluated documents were also provided by the authors of TAC-2011 [25]. This summarization will be used in comparison with our results. Our proposed method deals with English language only.

ROUGE [26] will be used to evaluate the performance of the proposed system. ROUGE package produces three numbers representing: precision (P), Recall (R) and F−score. They are formulated as follows.

$$p = \frac{\sum S_i}{\sum length\ S_j} \tag{13}$$

Where the S_i number of sentences occurring in both system and ideal summaries
Sj the number of sentences in the system summary.

$$R = \frac{\sum S_i}{\sum length\ S_k} \tag{14}$$

Where the S_i number of sentences occurring in both system and ideal summaries
S_k the number of sentences occurring in ideal summary.

$$F = \frac{(1+\beta^2)R*P}{R+\beta^2 P} \tag{15}$$

Where $\beta = \frac{P}{R}$

7. Experimental Results

Table 2- shows the results of our proposed MDS method and system summary that included in the TAC-2011 dataset [25] using ROUGE-1.

Table 2- Proposed MDS results

	Proposed MDS Results			System Results		
ID Number	Precision	Recall	F-Score	Precision	Recall	F-Score
ID1	0.45312	0.44421	0.44853	0.41253	0.40524	0.40776
ID2	0.52301	0.51232	0.51749	0.45655	0.46481	0.46062
ID3	0.49313	0.49131	0.49221	0.47909	0.43169	0.45404
ID4	0.48253	0.51324	0.49646	0.44966	0.44423	0.44691
ID5	0.52413	0.51023	0.51689	0.43513	0.41092	0.42243
ID6	0.51452	0.41321	0.44777	0.45122	0.35471	0.39617
ID7	0.44213	0.43214	0.43696	0.3953	0.39586	0.39547
ID8	0.43452	0.42341	0.42874	0.39265	0.38714	0.38985
ID9	0.39923	0.40212	0.40065	0.37726	0.38105	0.3791
ID10	0.5810	0.57203	0.57641	0.51806	0.52488	0.52141

As it's clear the results of the proposed method are better than the results of the peer summaries and that because of two reasons; the first reason the effect of the selected features which improves the performance of a TOPSIS method. The second reason most of the ATS methods may be affected by one feature that makes the sentence score high, while TOPSIS computes the effect of all features in the selected sentences.

8. Conclusions

The need for MDS increases with the rapid growth of information in the Internet. In this paper a method for MDS has been proposed which depends on TOPSIS. There are two important points in TOPSIS over other sentence ranking techniques. The first point TOPSIS depends on ranking the sentences according to the effect of all feature whereas in other method the effect of one feature may exceed the effect of other features which allows the sentence to take a high score. The second point is the weights of the features that are calculated mathematically using MLR to overcome the problem of assigning weight manually.

References

1. Alguliev , R. M., Aliguliyev, R. M., and Isazade, N. R. **2013**. Expert Systems with Applications Multiple documents summarization based on evolutionary optimization algorithm. *Expert Syst. Appl.*, **40**(5): 1675–1689.
2. Kumar ,R. and Chandrakal, D. **2016.** A survey on text summarization using optimization algorithm. *ELK Asia Pacific Journals,* **2**(1).
3. Huang, L., He, Y., Wei, F., and Li, W. **2010.** Modeling Document Summarization as Multi-objective Optimization. In Intelligent Information Technology and Security Informatics (IITSI), Third International Symposium on, pp:382-386. IEEE.

4. Song ,W., Cheon, L., Cheol, S., and Feng, X. **2011.** Expert Systems with Applications Fuzzy evolutionary optimization modeling and its applications to unsupervised categorization and extractive summarization. *Expert Syst. Appl.*, **38**(8): 9112–9121

5. Babar, S. A. and Patil, P. D. **2015.** Improving Performance of Text Summarization. *Procedia Comput. Sci.*, Icict 2014, pp: 354–363.

6. Song, W., Cheon, L., Cheol, S., and Feng, X. **2011.** Expert Systems with Applications Fuzzy evolutionary optimization modeling and its applications to unsupervised categorization and extractive summarization. *Expert Syst. Appl.*, **38**(8): 9112–9121.

7. Ferreira, R., De Souza, L., Dueire ,R., De Frana, G., Freitas, F., Cavalcant, G. D. C. i, Lima, R., Simske, S. J., and Favaro, L. **2013.** Assessing sentence scoring techniques for extractive text summarization. *Expert Syst. Appl.*, no1. May.

8. Megala, S. S. and Kavitha, A. **2014.** Feature Extraction Based Legal Document Summarization. *ijarcsms*, **2**(12): 346–352.

9. Binwahlan, M. S., Salim, N., and Suanmali, L. **.**Swarm Based Features Selection for Text Summarization. *IJCSN* , **9**(1): 175–179.

10. Abuobieda, A., Salim, N., Albaham, A. T., Osman, A. H., and Kumar Y. J. **2012.** Text summarization features selection method using pseudo genetic-based model. Proc. - 2012 Int. Conf. Inf. Retr. Knowl. Manag. CAMP, pp. 193–197.

11. Ghalehtaki, R., Khotanlou,H. and Esmaeilpour ,M. **2014.** A combinational method of fuzzy , particle swarm optimization and cellular learning automata for text summarization. IEEE conference, **15**(1).

12. Saleh**,** H. H., Kadhim, N. J. **2016.** Extractive Multi-Document Text Summarization Using Multi-Objective Evolutionary Algorithm Based Model. *Iraqi Journal of Science,* **57**(1C): 728-741.

13. Luo, W., Zhuang, F., He, Q., and Shi, Z. **2013.** Exploiting relevance , coverage , and novelty for query-focused multi-document summarization. *Knowledge-Based Syst.*, **46**: 33–42.

14. Porter stemming algorithm: http://www.tartarus.org/martin/PorterStemmer

15. JOHN, A. **2016.** Multi-Document Summarization System: Using Fuzzy Logic and Genetic Algorithm. *Int. J. Adv. Res. Eng. Technol*., **7**(1): 30 – 40 .

16. Satoshi, C. N. , Murata, M., Uchimoto, K., Utiyama, M. and Isahara, H. **2001.** Sentence Extraction System Assembling Multiple Evidence. Proc. 2nd NTCIR Work., pp. 319–324.

17. John, A. and Wilscy, D. M. **2013.** Random Forest Classifier Based Multi-Document Summarization System. *IEEE Recent Adv. Intell. Comput. Syst.* , pp. 31–36.

18. Chaterjee, S. and Hadi, A. **2006.** *Regression analysis by example.* A John Wiley & Sons, Inc., Publication Fourth Edition.

19. Timothy, Z. **2015.** *Multiple Regression and Beyond: An Introduction to Multiple Regression and Structural Equation Modeling*. By Routledge, Second edition published.

20. Hwang, C.L.; Yoon, K. **1981.** *Multiple Attribute Decision Making: Methods and Applications.* New York: Springer-Verlag.

21. Hwang, C. L., Lai, Y.J., Liu, T.Y. **1993.** A new approach for multiple objective decision making. *Computers and Operational Research*, **20**(8): 889-899.

22. Zavadskas, E., Zakarevicius, A. and Antucheviciene, J. **2006.** Evaluation of Ranking Accuracy in Multi-Criteria Decisions. *Informatica*, **17**(4): 601-618.

23. García-cascales, M. S. and Lamata, M. T. **2012.** On rank reversal and TOPSIS method. *Math. Comput. Model.*, **56**(5): 123–132.

24. Thanh, P. and van, P. **2016.** Project Success Evaluation Using TOPSIS Algorithm. *Journal of Engineering and Applied Sciences*, **8**: 1876-1879.

25. Giannakopoulos, G. and El-Haj, M. and Favre, B. and Litvak, M. and Steinberger, J. and Varma, V. **2011.** *TAC 2011 MultiLing pilot overview* . In: Text Analysis Conference (TAC) 2011, MultiLing Summarisation Pilot : TAC, Maryland, USA.

26. Lin, C.-Y. **2004**. ROUGE: a package for automatic evaluation summaries, in: Proceedings of the Workshop on Text Summarization Branches Out, Barcelona, Spain, July 25–26, pp. 74–81