



Multilingual Text Summarization using Deep Learning

Rana Talib Al Timimi¹, and Fatma Hassan Al Rubbiay²

¹ Department of Finance and Banking Sciences

² Department of Statistics

College of Economics and administration, Mustansiriyah University

Baghdad, Iraq

ABSTRACT

Along with the extreme expansion of big data and the vast development of the internet, making documentation of the huge internet information is the first interest for people. These online textual data led to information overload and redundancy. Multi-document summarization is one of the solutions to such an issue, used to extract the main ideas of the documents and put them into a short summary. Summarizing documents should not affect the major concepts and the meaning of the original text. This paper proposes a new method for multi-document summarization. The basic idea of the proposed method relied on six different features to be extracted of each sentence in the studied collection, these features must be language. A set of the feature vectors is introduced to Convolutional Neural Networks (CNNs) for classification as either summary or non-summary sentences. A graph of summary sentences was generated and assigned scores by the TextRank algorithm. The implemented system was evaluated on both English and Arabic versions of the dataset of the TAC-2011 MultiLing Pilot by using ROUGE metrics. The proposed method achieved an average F-measure 0.46079, 0.20664 using ROUGE-1 and ROUGE-2 respectively, for English documents, and achieved an average F-measure 0.45624, 0.30725 for Arabic documents.

Key Words: Summarization, TextRank, CNN, ROUGE, TAC-2011.

1. INTRODUCTION

With the great expansion of data communication strategies, e-documents were massively produced. This massive amount of information makes it more difficult to efficiently access the usable information [1]. This would be considered as the slowest process. Even when the best option was selected, the searcher might have a problem within understanding the validity and originality of the information [2]. One way to solve such problem of information overload is by generating summaries. Automatic Text Summarization (ATS) is a process of automatically creating a shorter version of a document or a set of documents by reducing the document(s) in length by selecting important information and discarding unimportant and redundant information, keeping the key elements and the basic meaning of information [3]. It provides readers with easier access to information without the labor of reading the source text. ATS has many applications in various related NLP processes such as headline generation, text classification, legal text summarization, and question answering [4].

Regarding the number of input documents to be summarized, the summary can be a single-document summarization (SDS), which produces the summary from a single document, or a multi-document summarization (MDS), which extracts the summary from a set of documents [5]. The determination of significant sentences from a single summary is a lot simpler, assuming that the order of the selected sentences maintains the order as in the original document, the summary still have a coherent document. One problem with MDS is determining what sentences to be selected from the documents and how to present it (in what order), coherence is a difficult problem to address moreover, the redundancy of information across documents is very important [6]. The repetition happens in light of the fact that the summarized documents share the same topics. The generated summary must not include any redundant information. On the other hand, SDS did not suffer from redundant information [7].

In general, summarization approaches are classified into two categories: extractive approach and abstractive approach. The extractive approach is designed to identify and select the most relevant sentences exactly as they appear in the original documents. Abstractive summarization, on the other hand is a technique in which the summary is created by generating novel sentences by either rephrasing or using the new words, instead of simply extracting the important sentence [8]. Abstractive summarization is more efficient than extractive summarization as it generates humanlike summaries, but it's more difficult to implement since it required deep NLP understanding techniques [9]. In addition; summarization can be classified based on the output summary as:

Generic or Query-Based. Generic text summarization distills the summarized text and presents a general sense of its contents. While Query-based text summarization is aimed at extracting essential information that answers the query from the original text. The answer is presented in a minimal, often predefined, number of words [10].

Summarization can be also classified based on the language into three categories: monolingual, multilingual, or cross-lingual. A summarization system is monolingual when the language of the source and target documents is the same. A summarization system is a multilingual when the source text is written in various languages such as Arabic, English, and French; the generated summary is also in these languages. Whereas a cross-lingual summarization is the task to produce a summary in one particular language for a source document in a different language [11]. Figure 1 shows the classification of ATS systems.

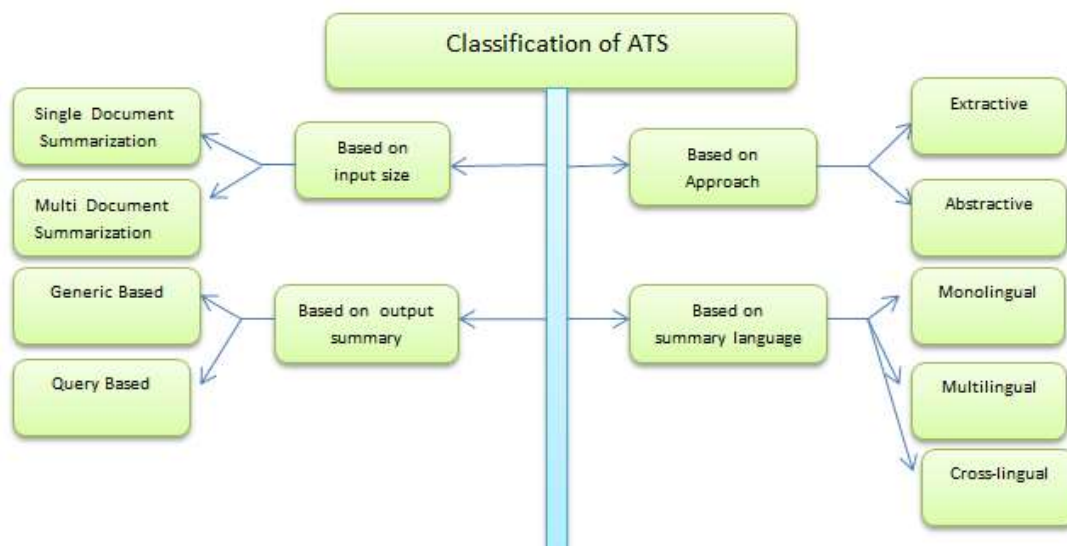


Figure 1: Classification of ATS Systems

The main contribution of this paper is the ability of the proposed method to deal with Arabic and English documents. The proposed method is a language-independent model which can be utilized for various languages. It does not require any extra external linguistic knowledge.

2. RELATED WORKS

Since the first method proposed by Luhn in 1958 [12], ATS has been commonly addressed by the research community with the proposal of various techniques, most of them based on monolingual summarization. Within the past two decades, a lot of works have focused on multilingual text summarization [13]. Radev et al. in 2004 developed MEAD MDS for English and Chinese languages. MEAD has been widely used in many summarization applications such mobile devices, search engine, and information retrieval. The method based on computing the cosine similarity of the feature vector with the centroid of the cluster. The SummBank and CSTBank corpora used to evaluate the system performance [14]. In 2006 Fung and Ngai proposed a method for multilingual MDS for both English and Chinese languages. The method based on segmented the input documents into many clusters using modified K-mean algorithm. The Hidden Markov Story Model (HMSM) used to train each cluster individually. The marked sentences output from HMSM introduced to the Naïve Bayes classifier for the classification purpose to embed into the created summary. TDT3 dataset used to evaluate the proposed method. The accuracy achieved 67.02 % for Chinese documents and 54.33 % for English documents [15]. In 2007 Patel et al proposed a method for SDS based on a language independent algorithm for English, Hindi, Gujarati and Urdu documents. The proposed method based on creating a feature vector for every sentence in the document collection. Sentences with highest weight selected to be in the summary. DUC dataset used for English language and several newspaper articles for other languages [16]. In 2010 Litvak et al. proposed a MUSE (MULTILINGUAL Sentence Extractor) for SDS that applied to English and Hebrew languages. The proposed method based on using genetic algorithm to compute the optimal weight for 31 statistical sentence scoring methods that are all language-independent and are based on either a vector or a graph representation of a document. DUC 2002 dataset used to evaluate the performance of English language and 50 news articles from Haaretz newspaper for Hebrew language [17].

In 2010 Kabadjov proposed NewsGist for multilingual MDS. The method based on computing the Singular value Decomposition (SVD). The proposed method consists of three main phases: interpretation, transformation and generation. The term-by-sentence created for the interpretation phase. Then the SVD algorithm applied for the transformation phase. In the last the best sentences score selected to include in the summary. The proposed algorithm applied to Europe Media Monitor that consists of large new articles in many languages [18]. In 2011 El-Haj et al. used a K-means clustering algorithm for MDS for both Arabic and English languages. The proposed method starts by selecting a number of sentences as the initial centroids. The algorithm then iteratively allocates all sentences to the nearest cluster, and recalculates the centroid of each cluster, until the centroids no longer change. The TAC-2011 multilingual dataset used to evaluate the proposed method [19]. In 2015 Giannakopoulos et al. proposed a multilingual text summarization for SDS and MDS. They present an overview of MultiLing 2015, a special session at SIGDial 2015 that help to introduce pilot tasks to promote research in summarizing human dialog in online fora and customer call centers [20].

In 2016 Singh proposed a method for SDS that deals with English and Hindi languages. The proposed method based on extracting eleven features from each sentence in the document collection. Unsupervised deep learning, restricted Boltzmann machine used to select the most relevant sentences based on the trained feature vector [21]. In 2019 Ouyang et al. proposed abstractive text summarization for Somali, Swahili, and Tagalog by using an automated translation of the New York Times dataset [22].

3. THEORETICAL BACKGROUND

In this section, a brief background to the CNN and TextRank are described.

3.1 CNN

Deep learning aids the learn of data representation in computational models which contained multiple processing layers by using multiple levels of abstraction [23]. Attracting interest across a variety of domains is considered as a major task of deep neural network. CNN is one of the dominant classes in computer vision field [24]. CNN is designed specially to learn features through backpropagation. Its components consist of three main layers: Convolutional layers, pooling layers, and fully connected layers. A convolution layer is a first layer of the CNN that apply a convolution operation to the input, passing the result to the next layer, converts all the input in its receptive field into a single value. A pooling layer is the next layer after the convolutional layer. Applied to the output of the convolutional layer separately to create a new set of the same number of pooled feature maps. The fully connected layer is a standard feed-forward neural network. It is a final straight line before the finish line where all the things are already evident [25].

CNN has many advantages including, It is easy to understand and fast to implement, Little dependence on preprocessing, decreasing the needs of human effort, and it works well both for supervised and unsupervised Learning [26]. In addition, CNN widely used in many applications such face recognition[27], scene labeling[28], image classification[29], and sentence classification [30].

3.2 Graph Approach

In recent years, the use of a graph technique for mining texts is a growing area of study. Its goal is to find novel and perceptive knowledge from information that is addressed as a graph. The utilization of this kind of graph methods on text documents has an extensive variety of applications like text clustering, word sense disambiguity, healthcare, web analysis, etc [31].

The graph can be either directed or undirected. The set of vertices (nodes) which are connected by directed edges, is known as a directed graph. While edges in an undirected graph are bidirectional, meaning no direction associated with them. Formally, the graph can be defined as $G=(E,V)$. Where V is a nonempty set of vertices, and E , a set of ordered pairs of distinct elements of V called edges where $V = (V_i, V_j \dots)$ and $E = \{e_1, e_2, e_3 \dots e_k\}$ and $e_i = (V_i, V_j)$ [32].

4. THE PROPOSED MDS MODEL

The goal of the ATS is to find a set of sentences that reflect the main content of the summarized documents. A good summary must include three important properties.

- Relevance: The selected sentences should comprise informative textual units that are important to the users.
- Redundancy: The generated summary must not contain redundant information.

- Length: summary should be bound to a specific length.

These three aspects jointly is the challenging task for the MDS. That is the reason the consideration of of relevant textual units depends not just on the properties of the actual units, yet in addition properties of each and every other textual unit in the summary.

4.1 Mathematical formulation of problem

Consider a document collection $D=\{d_1,d_2,\dots,d_n\}$ where n is the number of documents in D . Each d_i contains a set of sentences $S=\{s_1,s_2,\dots,s_p\}$ where p is the number of sentences in d_i . in this study, the document collection D is represented as the set of all sentences in the collection $D=\{s_1',s_2',\dots,s_N'\}$ where N is the total number of sentences in the document collection. The point to be reached is to get a set of $S_i \subset D$ that satisfy the three aspects of ATS. Through this MDS model, sentences are ranked and evaluated based on extracted features, then the most important sentences are selected from D to generate the summary.

4.2 Preprocessing and Feature Extraction

After applying the preprocessing step that includes sentence segmentation, tokenization, removing stop words and stemming, the features extraction steps begins, where the input document are represented as a vector of features. These features are the attributes that are used to represent summary sentences. Six features $F=\{f_1,f_2,\dots,f_6\}$ were extracted from each sentence in the document collection. The value of each feature ranges between 0 and 1, with values close to 1 indicating the high occurrence of the feature in the sentence, while values close to 0 indicating the low occurrence of the feature in the sentence. The extracted features are similar to [33] and [34] as explained below.

1. Sentence position: The location of sentences in the document is very important, where most instructive information is put in the front of the document. The importance of the sentences decreases gradually as the location of sentences moves to the end of the document. This features calculated as in Eq.1.

$$f1(s_i) = \begin{cases} 1 & \text{if } i \leq 3 \\ 1 - \frac{i-3}{i} & \text{if } i > 3 \end{cases} \quad (1)$$

Where i is the sentence position in the document.

2. Sentence length: This feature was computed based on disregarding very short and very long sentences. The length of the sentence is calculated as in Eq.2.

$$Al(S) = \frac{\max(\text{sentence_length}) + \min(\text{sentence_length})}{2} \quad (2)$$

$$f2(s_i) = \frac{|Al(S) - SL|}{\max(\text{sentence_length})} \quad (3)$$

Where $AL(S)$ is the average sentence length.

3. Numerical information: the presence of numerical information in the sentence increases its importance and may be considered as a selection choice into the summary. This feature is calculated as in Eq.4.

$$f3(s_i) = \frac{\text{num_count}(s_i)}{\text{length}(s_i)} \quad (4)$$

4. Similarity with the title: compute the overlap between the sentence and the title of the document. Eq.5 shows how this feature was computed.

$$f4(s_i) = \frac{S_i \cap \text{Title}}{\text{length}(\text{Title})} \quad (5)$$

5. Thematic words: The words that appear more frequently in the document seem to be more important. This feature can be computed by dividing the frequency of the top selected words by the highest word frequency as in Eq. 6.

$$f5(S_i) = \frac{\text{No.of thematic words in } S_i}{\text{Max number of Thematic}} \quad (6)$$

6. Sentence centrality: based on computing the similarity between the sentence (S_i) and all other sentences in the document as in Eq.7.

$$f6(S_i) = \frac{S_i \cap S_j}{S_i \cup S_j} \quad (7)$$

4.3 CNN as Sentence Classifier

Sentence classification could be applied using a simple CNN model which performs extremely well as shown in [30]. Supervised CNN is used to classify each sentence as either summary or non-summary sentence based on sentence existing or not in the ideal summary. The six extracted features from section 4.2 are fed to the CNN for the classification purpose. To learn the feature vector for each sentence and assign a label to it, CNN model is applied.

The implemented model is similar to that described in [30] and is composed of one convolutional layer, one max pooling layer and one fully connected layer. As known there are two important phases for any supervised learning: training and testing. The training phase contains input document collection and the ideal summaries. The input is represented as a set feature vector for every sentence in the documents. A label is given to the sentence based on computing ROUGE-1 between the input sentence and the ideal summaries. ROUGE-1 is a toolkit package used here to compute the unigrams similarity between input sentence and ideal summaries. The output of ROUGE-1 ranges from 0 to 1. A sentence is classified as a summary sentence when the output of ROUGE-1 exceeds a specified threshold. For the testing phase, each sentence will be used as input to the trained CNN model to be classified as summary or non-summary sentence, where the summary sentences pass to the next step of the proposed summarization model.

4.4 Sentences Scoring

Mainly there are two steps to score sentences, one is representing the sentences as a graph, the other is scoring graph node. During the first step, The sentences that classified as summary sentences from the CNN model are passed to the next step to build a graph. The graph consists of nodes and edges, where sentences are represented as nodes and the edges represent the relationship between nodes. ROUGE-1 used to compute the relationship between sentences. The relationship between sentences is key of constructing a graph. In addition, to show the significance of the nodes in the graph a similarity threshold θ is set, when an edge between two nodes is set if the similarity exceeds a θ value, otherwise the edge is discarded. This process helps to keep the consistency among the nodes in the graph by eliminating a connection when there is a small similarity between them.

After the first step a graph was obtained and the connections among nodes were established, then the TextRank algorithm [35] used to compute the score of every node in the graph. To decide the node importance within a graph, TextRank algorithm is used depending on a global recursively drawn information from the entire graph. The TextRank algorithm is formulated as in Eq.8.

$$TextRank(V_i) = (1 - d) + d * \sum_{V_j \in In(V_i)} \frac{W_{j,i}}{\sum_{V_k \in Out(V_j)} W_{j,k}} TextRank(V_j) \quad (8)$$

Where

- d is a damping factor.
- $In(V_i)$ is a set of input link that point to the node (V_i).
- $Out(V_j)$ is a set of output link from node (V_j).
- $W_{j,i}$ is the weight between node i and j calculated using ROUGE-1.

4.5 Sentences Selection

After scoring the sentences by the proposed method, most important sentences should be selected to form final summary. The selection is based on satisfying properties that described in section 4. The sentences are sorted in descending order from highest to the lowest based on the output of TextRank algorithm. The sentences are added one by one to the summary, taking into account the calculation of the similarity between the newly added sentence with the existing sentences to ensure non-repetition, and this process is repeated until a specific summary length is reached. ROUGE-1 used to calculate

the similarity between new selected sentence and the sentences that exists in the summary. The selected sentence is neglected when the similarity is greater than specific threshold, otherwise it added to the summary.

5. EXPERIMENTAL RESULTS

The evaluation of the proposed MDS was applied to the Arabic and English languages.

5.1 Dataset

The Text Analysis Conference TAC-2011 MultiLing Pilot dataset was used for MDS. TAC-2011 consists of seven languages (English, Arabic, Greek, Czech, French, Hindi, and Hebrew). The main documents were obtained from WikiNews website, which covers an assortment of news topics. The creation of the original dataset was started with English language, the dataset was translated to other languages. The dataset contains 10 topics, each topic consists of 10 documents, so a total of 100 documents available for each language as illustrated in table (2.1). Fluent speakers provide three model summaries for each topic in the dataset. There are 12 people cooperated in translating the English dataset to the Arabic dataset. In addition to the dataset the creator provides three peers (ideal) summaries for each topic, where these peer summaries are used for evaluation. For English language, there are 10 peer summaries ID1, ID2, ID3, ID4, ID5, ID6, ID7, ID8, ID9 and ID10, whereas there are 9 peer summaries ID1, ID2, ID3, ID4, ID6, ID7, ID8, ID9 and ID10 for Arabic language [36].

Table-1: TAC-2011 Dataset Description

Number of languages	7
Total number of documents in each language	100
Number of topics	10
Number of documents per topic	10
Source documents	WikiNews
Summary length	240-250 words

5.2 Evaluation metrics

Recall-Oriented Understudy for Gisting Evaluation (ROUGE) used to evaluate the proposed method [37]. It specifies the quality of a summary, which is automatically measured, by comparing its measured values with another summary(ideal) that created by humans. The basic idea based on computes the number of n-grams similar between the evaluate summary and the ideal-human summary. ROUGE-1 and ROUGE-2 used widely to evaluate automatic summary generation. They calculate the unigrams and bigrams overlap between each word present in the ideal summary and actual summary. The evaluation is calculated with respect to precision, recall and F-measure as in Eq.9, Eq.10 and Eq.11 respectively.

$$Precision = \frac{System\ Summary \cap ideal\ summary}{System\ Summary} \tag{9}$$

$$Recall = \frac{System\ Summary \cap ideal\ summary}{ideal\ Summary} \tag{10}$$

$$F - measure = \frac{2 * Precision * Recall}{Precision + Recall} \tag{11}$$

5.3 RESULTS AND DISCUSSION

The proposed MDS method applied for both English and Arabic languages. After the preprocessing and feature extraction the CNN model applied to classify sentences as summary or non-summary sentences. This process helps to reduce the number of unnecessary sentences in the second stage. It is well known that one of the problems with the graph is the large number of sentences that can increase the size of the graph and also affect the scores of the sentences. These unnecessary sentences may increase the score of some sentences that are less relevant to be included in the summary. Therefore, eliminating unnecessary sentences by applying the CNN model and using a threshold value to establish a connection between

nodes help to reduce this problem. The results are compared with the results of Ref. [19] that included within the TAC-2011 dataset. Figures 2 and 3 shows the results of the proposed method with Ref.[19] using ROUGE-1 and ROUGE-2 for English documents.

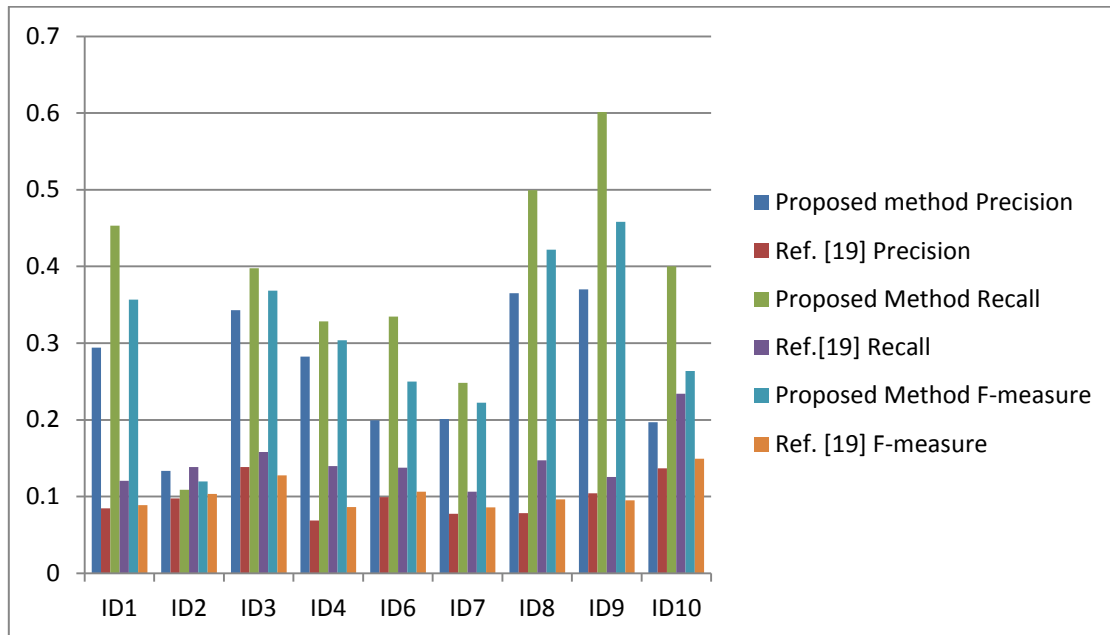


Figure 2: Comparison between Proposed Method and Ref. [19] using ROUGE-1 for English Documents

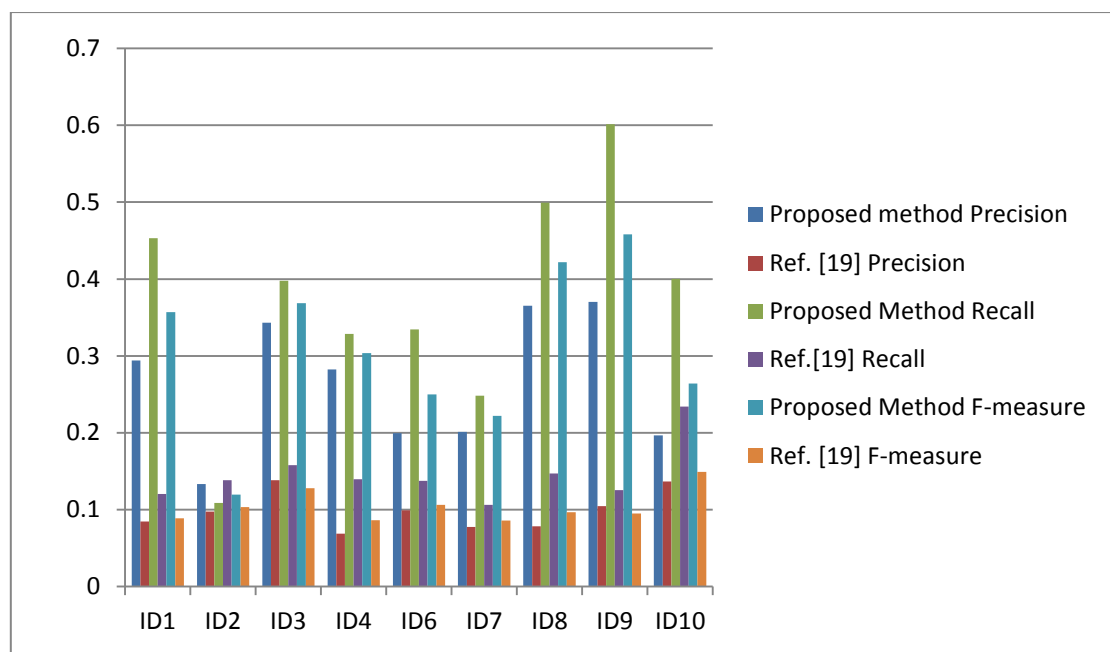


Figure 3: Comparison between Proposed Method and Ref. [19] using ROUGE-2 for English Documents

Table-2 shows the average comparison of precision, recall and f-measure for both proposed method and Ref.[19].

Table-2: Average Precision, Recall and F-Measure

		Precision	Recall	F-Measure
Proposed Method	ROUGE-1	0.38696	0.56999	0.46079
	ROUGE-2	0.18699	0.24576	0.20664
Ref.[19]	ROUGE-1	0.43674	0.42005	0.42737
	ROUGE-2	0.14950	0.14388	0.14634

As seen from Table-2 for ROUGE-1 the average recall and F-measure of the proposed method outperform Ref. [19] while it is less efficient in terms of precision. The Judgment between the recall and the precision is the F-measure that considers them both. The length of the system summary is the reason for these results. As is well known, the difference between precision and recall is the division by either the length of the system summary or the length of the ideal summary. Most system summaries are longer than ideal summaries, because the mechanism of creating system summary based on adding sentences to the summary until reaching specific length and without any change to the selected sentences. Sometimes the last sentence is too long that cause system summary length exceeds the specific length and that cause reducing the precision.

The effectiveness of the proposed system is shown through ROUGE-2 because it's closer to human summary than ROUGE-1. It clear from Table-2 the proposed method outperforms Ref. [19] in term of precision, recall and F-measure. The good performance obtained due to the efficiency of the proposed system, especially in terms of classification which reduced the number of sentences in the next step during sentences scoring.

The results of the proposed method and Ref.[19] are shown by figure 4 and 5 which use ROUGE-1 and ROUGE-2 for Arabic documents.

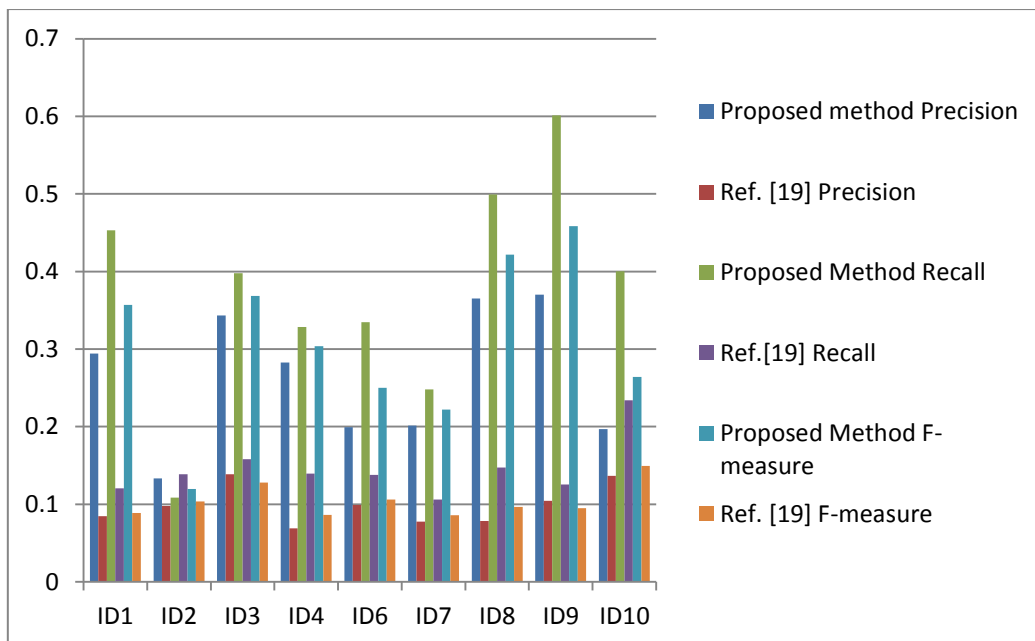


Figure 4: Comparison between Proposed Method and Ref. [19] using ROUGE-1 for Arabic Documents.

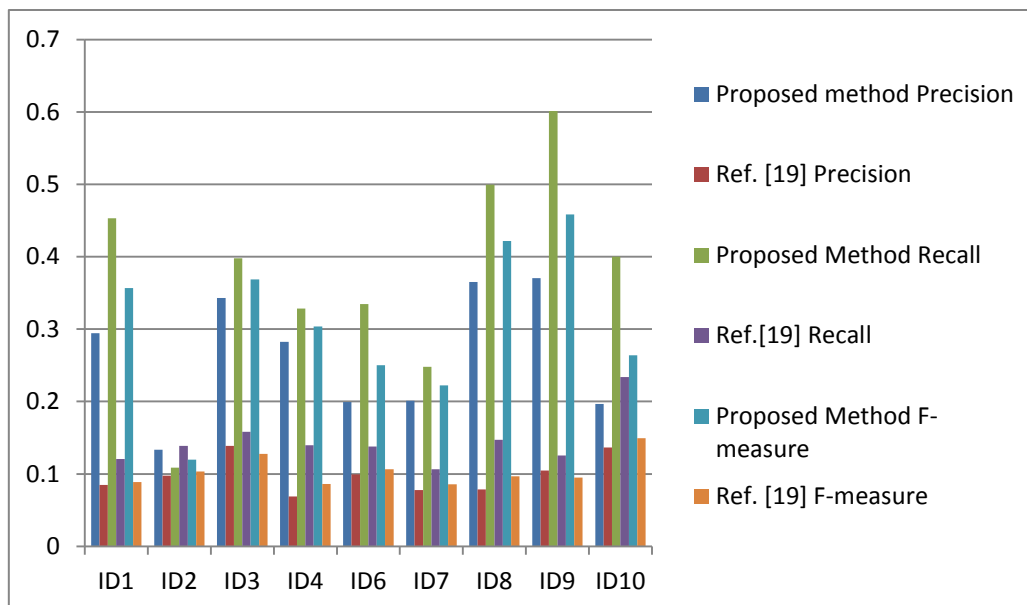


Figure 5: Comparison between Proposed Method and Ref. [19] using ROUGE-2 for Arabic Documents.

The average comparison of precision, recall and F-measure for the proposed method and Ref. [19] are shown in Table 3.

Table-3: Average Precision, Recall and F-Measure

		Precision	Recall	F-Measure
Proposed Method	ROUGE-1	0.38982	0.55065	0.45624
	ROUGE-2	0.26517	0.37464	0.30725
Ref.[19]	ROUGE-1	0.24517	0.35695	0.26591
	ROUGE-2	0.09851	0.14528	0.10436

It is clear from Table-3 that the proposed method outperform Ref. [19] in term of precision, recall, and F-measure. The combination of supervised method (CNN) and unsupervised method (graph) is the main reason for the well performance of the proposed method. The improvement effect over the TextRank comes from combining two methods, which indicates scoring the sentences from statistical features and sentences relationship comprehensively are more suitable to create a summary.

6. CONCLUSIONS

ATS is an influencing research aspect among the NLP field that helps produce concise information. The major concept of this paper is based on the combination of two methods for MDS purpose. At first, six independent features extracted for each sentence. That in turn classified as either summary or no-summary sentence using CNN. The graph was constructed for summary sentences and TextRank used to score the sentences. As it is known, dealing with the Arabic language is more difficult than the English language because of its absence of writing in capital letters, which makes it difficult to distinguish titles, acronyms, and abbreviations. Also, the derivation of the Arabic language, which makes morphological analysis not easy. Therefore, building a method that deals with Arabic and English documents not an easy task. The choice of features and the manner of sentence ranking is very important, since it required to implement for different languages.

For the future work, the author suggests studying the effect of using other classification method and sentence similarity measures. Also, discovering new features that are suitable to apply to more than one language

ACKNOWLEDGMENT

The authors would like to thank Mustansiriyah University (www.uomustansiriyah.edu.iq) Baghdad – Iraq for its support.

REFERENCES

1. Rasim M. A., Ramiz M. A. & Nijat R. I. (2012). DESAMC+DocSum: Differential evolution with self-adaptive mutation and crossover parameters for multi-document summarization, Knowledge-Based Systems , Vol. 36 , pp. 21–38.
2. Dutta, M. , Das, A. K. , Mallick, C. , Sarkar, A. , & Das, A. K. (2019). A graph based approach on extractive Summarization. In Emerging technologies in data mining and information security . Singapore: Springer . pp. 179–187.
3. Ermakova, L., Cossu, J. V., & Mothe, J. (2019). A survey on evaluation of summarization methods. Information Processing & Management, Vol. 56, pp.1794-1814.
4. Bidoki, M., Moosavi, M. R., & Fakhrahmad, M. (2020). A semantic approach to extractive multi-document summarization: Applying sentence expansion for tuning of conceptual densities. Information Processing & Management, Vol. 57, No.6.
5. Anand, D., & Wagh, R. (2019). Effective Deep Learning Approaches for Summarization of Legal Texts. Journal of King Saud University - Computer and Information Sciences. doi:10.1016/j.jksuci.2019.11.015.
6. Kanapala, A., Pal, S., & Pamula, R. (2019). Text summarization from legal documents: a survey. Artificial Intelligence Review, Vol. 51, No.3, pp. 371-402.
7. Ansamma, J., Premjith, P.S., Wilscy, M., (2017). Extractive multi-document ummarization using population-based multicriteria optimization. Expert Syst. Appl. <https://doi.org/10.1016/j.eswa.2017.05.075>.
8. Uçkan, T., & Karçı, A. (2020). Extractive multi-document text summarization based on graph independent sets. Egyptian Informatics Journal. doi:10.1016/j.eij.2019.12.002.
9. Bhargava, R., & Sharma, Y. (2020). Deep Extractive Text Summarization. Procedia Computer Science, Vol.167, pp. 138–146.
10. Bhargava, R., Sharma, Y., Sharma, G., (2016). Atssi: Abstractive text summarization using sentiment infusion. Procedia Computer Science Vol.89, pp. 404–411.
11. Wafaa S. El-Kassas , Cherif R. Salama , Ahmed A. Rafea, Hoda K. Mohamed ,(2021). Automatic text summarization: A comprehensive survey.Expert Systems with Applications. Vol.165.
12. Luhn, H. (1958).The automatic creation of literature abstracts. IBM Journal of research and development, Vol. 2, No. 2 , pp. 159-165.
13. Thomas, S., Paul-Alexis,D., Sylvain, L., Benjamin, P.& Jacopo, S.(2020), MLSUM: The Multilingual Summarization Corpus, [arXiv:2004.14900](https://arxiv.org/abs/2004.14900).
14. Radev, DR, Jing HY, Stys M, Tam D, (2004) Centroid-based summarization of multiple documents. Inf Process Manag ,Vol, pp.919–938.
15. Fung, P, Ngai G, (2006) One story, one flow: hidden Markov Story Models for multilingual multidocument summarization. ACM Trans Speech Lang Vol.3, pp.1–16. doi:10.1145/1149290.1151099.
16. Patel A, Siddiqui T, Tiwary US, (2007) A language independent approach to multilingual text summarization. In: Large scale semantic access to content (text, image, video, and sound), pp.123–132.
17. Litvak,M, Mark, L & Menahem, F.(2010), A new Approach to Improving Multilingual Summarization using a Genetic Algorithm, Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics, pp. 927–936.
18. Kabadjov M, Atkinson M, Steinberger J. (2010) NewsGist: a multilingual statistical news summarizer. Lecture notes in computer science, pp 591–594. doi:10.1007/978-3-642-15939-8_40.
19. El-Haj M, Kruschwitz U, Fox C. (2011) University of essex at the tac 2011 multilingual summarisation pilot. In: Proceedings of the text analysis conference (TAC) 2011, MultiLing Summarisation Pilot, Maryland, USA.
20. Giannakopoulos,G, Jeff ,K, John, C,Josef ,S, Benoit ,F, Mijail, K, Udo, K, & Massimo P.,(2015). Multiling 2015: multilingual summarization of single and multi-documents, on-line fora, and call-center conversations. In Proceedings of the 16th Annual Meeting of the Special Interest Group on Discourse and Dialogue, pp. 270–274.
21. Singh,S, Ajai, K., Abhilasha M. & Shikha S.,(2016), Bilingual Automatic Text Summarization Using Unsupervised Deep Learning, International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT), pp.1195-1200.

22. Ouyang, J., Boya, S. & Kathleen ,K.(2019). A robust abstractive system for cross-lingual summarization. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Vol 1, pp. 2025–2031.
23. Yann ,L., Yoshua, B. & Geoffrey, H.,(2015). Deep learning. Macmillan Publishers,Vol. 521, pp.436-444.
24. Yamashita, R., Nishio, M., Do, R. K. G., & Togashi, K. (2018). Convolutional neural networks: an overview and application in radiology. Insights into Imaging. doi:10.1007/s13244-018-0639-9.
25. Ashwin, B., Maithili, B., Pranav, G. & Rohan, C.(2016). Applications of convolutional Neural Networks,Vol 7, No. 5, pp. 2206-2215.
26. Wajdi, H. & Norjihan,B.(2019). Convolutional Neural Network based for Automatic Text Summarization, International Journal of Advanced Computer Science and Applications, Vol. 10, No. 4.
27. Lawrence, S.(1997). Face recognition: A convolutional neuralnetwork approach. IEEE transactions on neural networks , pp. 98-113.
28. Farabet, C., Camille, C., Laurent, N.& Yann, L.(2013) Learning hierarchical features for scene labeling.IEEE transactions on pattern analysis and machine intelligence , pp.1915-1929.
29. Chen, L., et al.(2014) Semantic image segmentation with deep convolutional nets and fully connected crfs."arXiv preprint arXiv:.1412.7062.
30. Kim,Y.,(2014). Convolutional neural networks for sentence classification, arXiv preprint arXiv:1408.5882.
31. Kamal, S. & Khushbu, S. (2015), Improving Graph Based Multidocument Text summarization Using an Enhanced Sentence Similarity Measure, IEEE 2nd International Conference on Recent Trends in Information Systems, pp. 395-365.
32. SAKIRA, K. Siti et al.(2018), Graph-based Representation for Sentence Similarity Measure : A Comparative Analysis. International Journal of Engineering & Technology, vol. 7, No. 2, pp. 32-35.
33. Oliveira, H., Ferreira, R., Lima, R., Lins, R. D., Freitas, F., Riss, M., & Simske, S. J. (2016). Assessing shallow sentence scoring techniques and combinations for single and multi-document summarization. Expert Systems with Applications,Vol. 65, pp.68–86.
34. Mao X, Yang H, Huang S, Liu Y, Li R.,(2019), Extractive summarization using supervised and unsupervised learning. Expert Syst vol.133,pp.173–81.
35. Mihalcea R & Tarau P.(2004), Textrank: Bringing order into text. In: Proceedings of the conference on empiricalmethods in natural language processing. pp. 404–11.
36. Giannakopoulos,G. El-Haj M. Favre,B. Litvak, M. Steinberger, Josef. & Varma,V. (2011). "TAC11 MultiLing Pilot Overview". TAC2011 Workshop. Presented at the TAC 2011 Gaithersburg,MD, U.S.A.
37. Lin, Chin-Yew (2004). ROUGE: A Package for Automatic Evaluation of summaries. In Proceedings of the Workshop on Text Summarization Branches Out (WAS 2004).