Automatic Arabic Text Summarization Based on Fuzzy Logic

Lamees M. Al Qassem
The Emirates ICT Innovation Center
Electrical and Computer Engineering Department,
Khalifa University,
Abu Dhabi, UAE
lamees.alqassem@ku.ac.ae

Di Wanga
The Emirates ICT Innovation Center
Khalifa University,
Abu Dhabi, UAE.
di.wang@ku.ac.ae

Hassan Barada
Electrical and Computer
Engineering Department,
Khalifa University,
Abu Dhabi, UAE
hassan.barada@ku.ac.ae

Ahmad Al Rubaiea, Nawaf Al Moosaa
The Emirates ICT Innovation Center
Khalifa University,
Abu Dhabi, UAE.
{ahmad.al-rubaie,nawaf.almoosa}@ku.ac.ae

Abstract

The unprecedented growth in the amount of online information available in many languages to users and businesses, including news articles and social media, has made it difficult and time consuming for users to identify and consume sought after content. Hence, automatic text summarization for various languages to generate accurate and relevant summaries from the huge amount of information available is essential nowadays. Techniques and methodologies for automatic Arabic text summarization are still immature due to the inherent complexity of the Arabic language in terms of both structure and morphology. This work attempts to improve the performance of Arabic text summarization. We propose a new Arabic text summarization approach based on a new noun extraction method and fuzzy logic. The proposed summarizer is evaluated using EASC corpus and benchmarked against popular state of the art Arabic text summarization systems. The results indicate that our proposed Fuzzy logic approach with noun extraction outperforms existing systems.

1 Introduction

In the recent two decades, the exponential growth in the amount of information like email, online news articles, reports, social media content and memos, introduced new challenges and made it harder for users to sift through and extract the key information they need. Hence, a smart system that can automatically identify important information from vast amount of data and generate concise summaries from these identified data is highly demanded nowadays. Automatic accurate text summarization is the key to addressing this challenge. Text summarization is the process of conveying important information from the original text source(s). The summary is typically no longer than half of the original text(s) and usually significantly less than that (Das and Martins, 2007). Techniques for automatic text summarization for widely-used and relatively simple-grammar languages such as English are mature. However, little work has been done for Arabic summarization (Al Qassem et al., 2017) due to the complexity of the language in terms of both structure and morphology. Nevertheless, Arabic summarization systems are highly needed nowadays. There are more than 300 million Arabic speakers in the world, and Arabic is an official language in the United Nations (Nenkova et al., 2011) and 22 other countries (Al-Shalabi et al., 2009). Therefore, researchers are working on improving Arabic text summarization methods and developing real world systems. A smart system is needed to automatically generate summaries from Arabic texts and deliver these summaries to the user, either directly or on-demand. The generated summaries need to be coherent, readable, grammatically correct, and comprise the key information of the original texts. This requires an in-depth study to achieve better pre-processing for Arabic text and a better methodology to extract the main information and generate a more accurate summary.

In this work, we propose and develop a smart

Arabic summarization system with better accuracy than the current state of the art. The system has been applied to generate summaries from online news in real time and delivers the summary instantly to the right users who just need it. The paper is organized as follows. The state of art for text summarization (mainly English and Arabic summarization systems) is discussed in Section 2. Our proposed summarization system is described in Section 3. The evaluation and comparison results are explained in Section 4. Our conclusion is given in Section 5.

2 Related Work

The first automatic summarization system was proposed by (Luhn, 1958). Luhn came up with the assumption that says the more frequent the word appears in the text the more important it is; excluding the very common words (called stop words). Ten years later, (Edmundson, 1969) expanded Luhns work by adding more features, such as resemblance to the title feature (the vocabulary overlap between the title and the sentence) and the position feature (the relevant position of the sentence within the text). The results showed that the word frequency is set to be the least important feature. It is important to note that the author assigned weights to the features subjectively; thus, these assigned weights could be imprecise and uncertain. In 1995, (Kupiec et al., 1995) developed a trainable document summarizer to automatically train the weights of the features using a corpus instead of defining the weights subjectively. The evaluation results agreed with Edmundsons results (Edmundson, 1969). As a conclusion, both work claimed that the best combination of features is made of the position feature, key word feature, and the title feature.

Although a lot of work has been done for text summarization for English, the work for Arabic summarization is very recent and limited. Lakhas (Douzidia and Lapalme, 2004) considered one of the first known Arabic text summarization system that was evaluated and compared to English systems. The system produces a summary of size of 10 words only and translates it to English. The authors claimed that the translation process is the reason for the bad evaluation scores. Using TF-IDF (term frequency-inverse document frequency) as the main feature to score sentences is a common method in Arabic summarization systems.

TF-IDF is the ratio of the frequency of a term in a document over its frequency in a corpus. TF-IDF is a good indicator of the importance of a word in a document and a topic, and hence highlights the importance of the corresponding sentence. (Haboush et al., 2012) used TF-IDF on clustered word roots and obtained competitive accuracy. (Al-Radaideh and Afif, 2009) developed a system that focuses on the inner product between TF in a sentence and the document frequency DF for each extracted noun. ACBTSS and AQBTSS (El-Haj et al., 2009) are two most recent systems that used TF-IDF with Vector Space Model (VSM). Semantic connectedness among sentences and documents is another important factor when generating summaries with minimum redundancy. LCEAS system (Sarmini, 2015) used lexical cohesion to identify important topics and text entailment to remove redundancy. Their system outperformed (Haboush et al., 2012), (Al-Radaideh and Afif, 2009), Sakhr, AQBTSS (El-Haj et al., 2009), Gen-Summ (El-Haj et al., 2010) and LSA-Summ (El-Haj et al., 2010), by containing more significant sentences and less redundancy. In recent work, more features/indicators are researched to represent the importance of sentences. Therefore, deciding which features to use and the weights for these features become a hard task and more research is needed. Some researchers followed the machine learning approach and modeled the summarization process as a classification problem (i.e. the sentences are classified as summary and non-summary sentences). The work in (Boudabous et al., 2010) and (Belkebir and Guessoum, 2015) includes examples of systems that followed the machine learning approach. In (Boudabous et al., 2010), SVM (Support Vector Machine) was used to classify the sentences using 15 features. In (Belkebir and Guessoum, 2015) an Arabic summarizer was proposed using AdaBoost. Machine learning approaches give researchers the ability to efficiently utilize a large number of features in the scoring process, which is desirable. Using fuzzy logic in text summarization is a very recent approach in English text summarization (Yadav and Meena, 2016). (Suanmali et al., 2009) used the fuzzy logic approach to select the sentences based on eight features. The system was compared to a baseline summarizer that generates summaries by selecting the first 200 words in the input document and MS word 2007 summa-
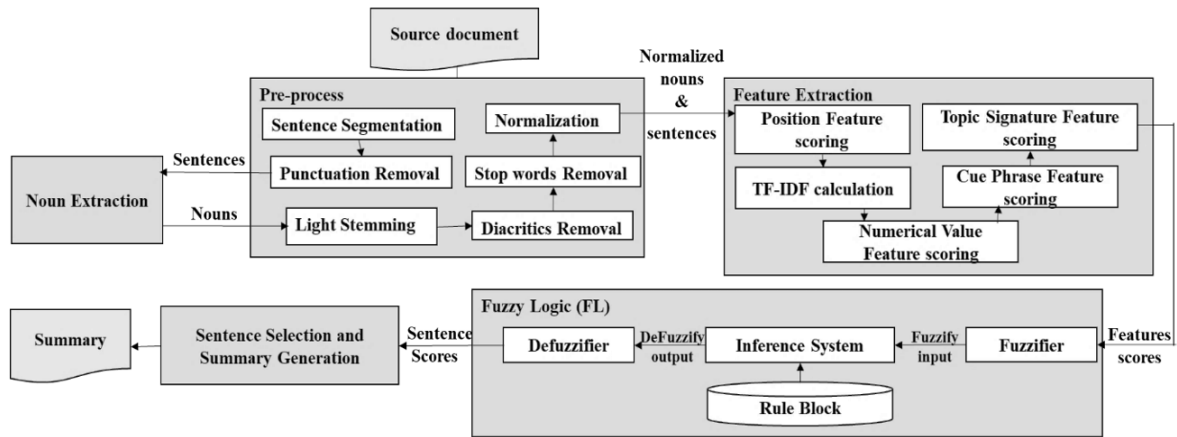
Figure 1: System Architecture.

rizer. The results showed that their proposed approach outperformed the baseline summarizer and MS word 2007 summarizer. (Yadav and Meena, 2016) used fuzzy logic along with WordNet synonyms and bushy path, a graph-based method, to improve the performance of extractive text summarization system. The WordNet synonyms is used for the semantic similarity of the text; bushy path is used for the relationship between different parts of the text; finally, fuzzy logic is used to solve the issue of uncertainty and vagueness related to the weights for different features of the sentences. The system generated three summaries from the three approaches and then selected the sentences that appeared in all summaries to form the final summary. The three approaches were evaluated and compared against the proposed approach using ROUGE-1 and ROUGE-2. The results showed that the proposed approach outperformed the other three approaches. In addition, the evaluation results showed that the fuzzy logic approach outperformed the bushy path and the WordNet synonyms methods. (Sarmini, 2015) proposed an Arabic text summarizer based on fuzzy logic and genetic algorithm. The genetic algorithm was used to select the optimal member functions of the selected features. The fuzzy system is used to score the sentences. To sum up, all reviewed systems claimed that using fuzzy logic improved the performance of the summarization systems and the quality of the summaries. Fuzzy logic approaches handled the issues related to the uncertainty, imprecision and vagueness of determining the importance of different features using machine learning approaches, leading to better summaries.

3 Proposed Fuzzy Logic Arabic Summarizer

Condensing all the discussions and comparisons in the literature review, we propose an Arabic summarization system with five main components (pre-processing, noun extraction, features extraction, fuzzy logic, sentence selection) to generate the final summary. Figure 1 shows the five components in the proposed system. The first two steps in the proposed system are Pre-processing and Noun extraction. Pre-processing prepares the text before sentences are further treated and summarized. Noun extraction extracts the nouns from the text output of the pre-processor. From the state of art noun words are considered to carry important information than other words (Al Qassem et al., 2017; Al-Radaideh and Afif, 2009). A noun is any word representing an idea, a thing or a person. To have a good summary, we need to make sure all sentences representing the main ideas are selected. To assure this, all nouns in the text should be processed and evaluated. The importance of a sentence will then be scored by the extracted nouns only. Furthermore, using nouns only will reduce noise and increase efficiency by avoiding unnecessary processing. In our previous work, we proposed a linguistic-rule-based noun extraction system (Al Qassem et al., 2018) that extracts nouns according to Arabic grammar rules. The system is evaluated against the widely used Stanford Arabic Part of Speech (POS) tagger (Stanford Log-linear Part-Of-Speech Tagger, n.d.). The results show that the proposed method is more efficient when achieving comparable benchmark accuracies. The details of our proposed Arabic noun extraction

method has been explained in our previous paper (Al Qassem et al., 2018) and will not be repeated here due to the size limit of the paper. After that the feature extraction module extracts key features (sentence position, TF-IDF, cue phrase, topic signature and numerical data) representing the importance of the sentences. Finally, the extracted features/scores are input into the fuzzy logic module to generate the final scores of the sentences. The sentence' score indicates how important a sentence is within the whole article. The sentences with the highest scores are selected to form the final summary. In our system we used five features based on the discussion and experimental results from the state of art (Ferreira et al., 2013; Fattah and Ren, 2009); they are: (1) Sentence position: this is just the position of the sentence within the full text; (2) TF-IDF: it is calculated for the extracted nouns only as a feature that indicates the importance accumulation of the extracted nouns in a sentence; (3) Cue phrases: they are phrases that give a good indication about the content of this sentence such as in conclusion, the most important ... etc., and defined as positive cue phrases (Ferreira et al., 2013). (Haboush et al., 2012) claimed that the existence of these cue phrases increases the probability for a sentence to be selected. On the contrary, there are list of phrases that give a detailed explanation or indicates redundant information like in other words and for example. These phrases are called negative cue phrases (Fattah and Ren, 2009). In our system, we use both types of cue phrases to either increase or decrease the importance score of the sentences. The two other features are: (4) Topic Signature: each topic has a list of topic signature words used across all documents within this topic but not frequently used across other topics(the score of the sentence that contains topic signature words is supposed to increase); and finally (5) Numerical Data: sentences that contain numbers are more likely to be added to the summary because numbers refer to important information like money transaction, dates, address ... etc. (Ferreira et al., 2013; Fattah and Ren, 2009). The final score of the sentences is calculated by combining all the features. The linear combination of all features (feature-weight equation) is usually used for the final score. The main challenge in this step is assigning a weight for each feature. As discussed previously, not all features are equally important and different fea-

tures should be given different weights representing their importance and contribution to generate a high quality summary. Therefore, we use fuzzy logic. At this stage the features extracted from the sentence are inputs to the fuzzy logic system, and the sentence final score is the output. According to (Hüllermeier, 2011), fuzzy logic can contribute in solving issues related to uncertainty, vagueness, ambiguity, and imprecision that result from incomplete and imprecise information. Fuzzy logic provides the ability to map rules using concept (e.g. long vs short, big vs small) rather than numbers (numerical data). Furthermore, representing gradual concepts is a key feature of fuzzy logic compared with machine learning that failed to do so (Hüllermeier, 2011). Fuzzy logic is transparent, data-driven and makes use of available expert knowledge (for model initialization) to generate a robust model. It is considered an approximate reasoning solution that can be initialized from expert knowledge and optimized from data with very strong reasoning capabilities (Megala et al., 2014). Finally, the sentences with highest scores from the fuzzy logic system are selected to form the summary. The sentences in the summary are ordered by their original position in the article. The proposed Arabic summarization system can generate different sizes of the summaries based on user choice. Our observations and evaluation of the generated model are aligned with our hypothesis, in that the first few sentences represent the main ideas. This is expected in news articles that tend to be relatively short; important words repeat more frequently within the text and cue phrases are used to attract the attention of the reader.

4 System Evaluation

Evaluating an Arabic text summarizer is a challenging task due to the lack of gold standard corpora and the different measures used in assessing summarization systems. We have therefore, decided to choose the corpus and evaluation metrics that are used by most benchmark systems in the literature, to provide as objective comparison as possible. Based on this approach, we found that ROUGE-N (N=1and 2) with EASC corpus (El-Haj et al., 2010) are used by many recent systems. ROUGE correlates well with human judgement for single-document summarization tasks. In addition, the correlation increases by using multiple references. This gives an advantage for EASC
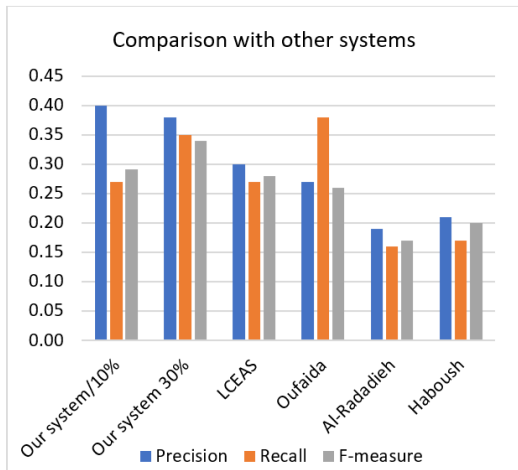
Figure 2: Performance comparison with other systems in the literature.

System	Average Recall ROUGE-		Average Precision ROUGE-		Average F-measure ROUGE-	
	1	2	1	2	1	2
Oufaida (2014)	0.41	0.38	0.37	0.27	0.36	0.26
Our Sys 10%	0.34	0.27	0.51	0.40	0.37	0.29
Our Sys 30%	0.45	0.35	0.48	0.38	0.44	0.34

Table 1: ROUGE evaluation results for our system and Oufaida system.

corpus as each document has five reference summaries.

We compared our system with the state of art systems (LCEAS 2015 (Al-Khawaldeh and Samawi, 2015), (Al-Radaideh and Afif, 2009), (Haboush et al., 2012) and (Oufaida et al., 2014)). Figure 2 below illustrates the comparison of ROUGE-2 results for our summarizer and the other systems. The 30% summary size represents the reference summaries that are neither too long nor too short, the 10% summary size represents the shortest summary. As shown in the Figure, the ROUGE-2 scores for LCEAS (Al-Khawaldeh and Samawi, 2015), (Al-Radaideh and Afif, 2009) and (Haboush et al., 2012) are less than 0.3 for recall and less than 0.2 for precision, where the best scores were obtained by LCEAS. The F-measure for LCEAS system is approximately 0.22. To compare our system against these three systems, we use the ROUGE results when the summary size is 10% from the proposed Arabic summarization system (the smallest possible size). The average ROUGE-2 recall, precision and F-measure scores for our system are 0.27, 0.40 and 0.29, respectively. Our system outperformed the three systems in F-measure scores despite the fact that these systems were compared against our worst-case results.

Furthermore, LCEAS was compared against Sakhr, AQBTSS (El-Haj et al., 2009), Gen-Summ and LSA-Summ (El-Haj et al., 2010). The authors claimed that LCEAS outperformed all these three systems. Since our fuzzy logic summarizer outperformed LCEAS, it is our logical assumption that our system will outperform these three systems too.

For (Oufaida et al., 2014), the system was evaluated using ROUGE-N (N=1 and 2) and EASC corpus.

The generated summary size of a document is equal to its reference summary. The EASC has five summaries per article. Consequently, the system generated five summaries per document and computed their average ROUGE-N scores. The system ROUGE-1 and ROUGE-2 scores are shown in Table 1. We compared our system against Oufaida using the ROUGE-results obtained for the summary sizes 30% and 10%, which are the percentages used in the state of art methods for our fair comparison. In the real world application, this percentage can be changed and adjusted by users based on the requirements. The less the percentage is, the more concise the summary is but some information might be missed. On the contrary, summary with higher percentage of length provides more information (sentences) but the summary takes more human beings time to read (hence not a very efficient summary). According to both summary sizes 10% and 30% and ROUGE-N (N=1,2) results, our system outperformed Oufaidas.

5 Conclusion

Due to the increase in the amount of information available online, consuming a broad range of relevant, concise but important information has become a laborious task. Automatic text summarization methods are put forward to address this problem. Text summarization for English is advanced and many approaches have been studied and evaluated. However, this field is still in its early stages for the Arabic language. In this paper, we discussed different text summarization ap-

proaches and methodologies and proposed our approach by using fuzzy logic for a more accurate and efficient Arabic summarization system. Fuzzy logic is still very recent in English summarization and showed improvement in the quality of the generated summaries. We compared our summarizer against five state of the art Arabic text summarizers that reported good results. The results showed that fuzzy logic improved the performance of the summarization system. The system is able to create very short summaries containing the most important ideas, and performed better than five state of art Arabic summarization systems. The future work might be looking into better Arabic preprocessing, (e.g. http://arabicnlp.pro/alp/) for more accurate Arabic summarizer.

References

F Al-Khawaldeh and V Samawi. 2015. Lexical cohesion and entailment based segmentation for arabic text summarization (lceas). *The World of Computer Science and Information Technology Journal (WSCIT)*, 5(3):51–60.

Lamees Al Qassem, Di Wang, and Hassan Barada. 2018. Noun extraction tool for anlp applications. In *2018 IEEE 12th International Conference on Semantic Computing (ICSC)*, pages 308–309. IEEE.

Lamees Mahmoud Al Qassem, Di Wang, Zaid Al Mahmoud, Hassan Barada, Ahmad Al-Rubaie, and Nawaf I Almoosa. 2017. Automatic arabic summarization: a survey of methodologies and systems. *Procedia Computer Science*, 117:10–18.

Q Al-Radaideh and Mohammad Afif. 2009. Arabic text summarization using aggregate similarity. In *International Arab conference on information technology (ACIT2009), Yemen*.

Riyad Al-Shalabi, Ghassan Kanaan, Bashar Al-Sarayreh, Khalid Khanfar, Ali Al-Ghonmein, Hamed Talhouni, Salem Al-Azazmeh, et al. 2009. Proper noun extracting algorithm for arabic language. In *International conference on IT, Thailand*.

Riadh Belkebir and Ahmed Guessoum. 2015. A supervised approach to arabic text summarization using adaboost. In *New Contributions in Information Systems and Technologies*, pages 227–236. Springer.

Mohamed Mahdi Boudabous, Mohamed Hédi Maaloul, and Lamia Hadrich Belguith. 2010. Digital learning for summarizing arabic documents. In *International Conference on Natural Language Processing*, pages 79–84. Springer.

Dipanjan Das and André FT Martins. 2007. A survey on automatic text summarization. *Literature Survey for the Language and Statistics II course at CMU*, 4(192-195):57.

Fouad Soufiane Douzidia and Guy Lapalme. 2004. Lakhas, an arabic summarization system. *Proceedings of DUC2004*.

Harold P Edmundson. 1969. New methods in automatic extracting. *Journal of the ACM (JACM)*, 16(2):264–285.

Mahmoud El-Haj, Udo Kruschwitz, and Chris Fox. 2009. Experimenting with automatic text summarisation for arabic. In *Language and Technology Conference*, pages 490–499. Springer.

Mahmoud El-Haj, Udo Kruschwitz, and Chris Fox. 2010. Using mechanical turk to create a corpus of arabic summaries.

Mohamed Abdel Fattah and Fuji Ren. 2009. Ga, mr, ffnn, pnn and gmm based models for automatic text summarization. *Computer Speech & Language*, 23(1):126–144.

Rafael Ferreira, Luciano de Souza Cabral, Rafael Dueire Lins, Gabriel Pereira e Silva, Fred Freitas, George DC Cavalcanti, Rinaldo Lima, Steven J Simske, and Luciano Favaro. 2013. Assessing sentence scoring techniques for extractive text summarization. *Expert systems with applications*, 40(14):5755–5764.

Ahmad Haboush, Maryam Al-Zoubi, Ahmad Momani, and Motassem Tarazi. 2012. Arabic text summarization model using clustering techniques. *World of Computer Science and Information Technology Journal (WCSIT) ISSN*, pages 2221–0741.

Eyke Hüllermeier. 2011. Fuzzy sets in machine learning and data mining. *Applied Soft Computing*, 11(2):1493–1505.

J Kupiec, J Pedersen, and F Chen. 1995. A trainable document summarizer. dans les actes de acm special interest group on information retrieval (sigir), 68–73.

Hans Peter Luhn. 1958. The automatic creation of literature abstracts. *IBM Journal of research and development*, 2(2):159–165.

S Santhana Megala, A Kavitha, and A Marimuthu. 2014. Enriching text summarization using fuzzy logic. *International Journal of Computer Science and Information Technologies*, 5(1):863–867.

Ani Nenkova, Kathleen McKeown, et al. 2011. Automatic summarization. *Foundations and Trends® in Information Retrieval*, 5(2–3):103–233.

Houda Oufaida, Omar Nouali, and Philippe Blache. 2014. Minimum redundancy and maximum relevance for single and multi-document arabic text summarization. *Journal of King Saud University-Computer and Information Sciences*, 26(4):450–461.

Mohamad Sarmini. 2015. Design and implementation of hybrid syntactic-fuzzy genetic system for extractive arabic text summarization. *International Journal of Computer Science and Engineering in Arabic*, 6:69–98.

Ladda Suanmali, Naomie Salim, and Mohammed Salem Binwahlan. 2009. Feature-based sentence extraction using fuzzy inference rules. In *2009 International Conference on Signal Processing Systems*, pages 511–515. IEEE.

Jyoti Yadav and Yogesh Kumar Meena. 2016. Use of fuzzy logic and wordnet for improving performance of extractive automatic text summarization. In *2016 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, pages 2071–2077. IEEE.