International Journal of Academic Information Systems Research (IJAISR)
ISSN: 2643-9026
Vol. 4, Issue 8, August – 2020, Pages: 6-9

Arabic Text Summarization Using AraBERT Model Using Extractive Text Summarization Approach

Abdullah M. Abu Nada[1], Eman Alajrami[2], Ahemd A. Al-Saqqa[3], Samy S. Abu-Naser[4]

[1,3] IT Unite, University of Palestine, Gaza, Palestine
[2] IT Department, University of Palestine , Gaza, Palestine
[4] IT Department, Al-Azhar University, Gaza, Palestine
Email: e.alajrami@up.edu.ps[1], a.abunada@up.edu.ps[2], ah.saqqa @up.edu.ps[3], abunaser@alazhar.edu.ps[4]

Abstract*—recently, after the life of the individual changed and became more crowded with all the concerns of life, and with the diversity and the increasing of sources of knowledge on the Internet, it became difficult for us to read large texts and articles, so we are looking for the summaries of these texts before deciding dive deeply in reading. For this reason, it became urgent to provide tools to fulfill this function by extracting basic information while preserving the essence of the text. In this study, we proposed an extractive Arabic text summarizer based on a general-purpose architecture for Natural Language Generation (NLG) and Natural Language Understanding (NLU) like (AraBERT, BERT, XLNet, XLM, etc.) to summarize the Arabic document by evaluating and extracting the most important sentences at this document. Then, using the Rouge measure and human evaluation, we compared the efficiency between the proposed and other solutions to recommend what the best one we can use to summarize Arabic text and put our hands-on weak points to open the way for researchers to improve the approaches.*

Keywords— Text Summarization, Arabic language, Extractive Summarization, Transformers, AraBERT Model

I. Introduction

As the textual information available online expands rapidly, it becomes difficult for the reader to read a large amount of text to know which of these texts are useful. Therefore, it became necessary for researchers to research at the Automatic text summarization field.

A summary also knew an abstract, "is a short version of text which is generated from one or more texts and covers the main information in the original texts and not to exceed the half length of the original text" [1].

Automatic text summarization is a process of generating a coherent, fluent and meaningful summary by covering the most important information as much as possible. Recently, may methods and techniques have been proposed for automatic text summarization and applied widely in various fields. For example, in social media marketing, it used to compress the text like blogs, e-books, whitepapers, etc., to adapt it to be shareable on social media [2].

This task can be done by two approaches, the first one is extractive summarization it's based on extraction the core and most informative sentences from the original text. On the other hand, abstractive summarization is generating new sentences with a linguistic structure that might be different and these sentences might not exist in the original text.

This field is rich of scientific research, but unfortunately, most of this research is based on the English language texts, and there is a lack of research on summarizing the Arabic texts. This is due to several reasons, first one is the complexity of Arabic language like grammatical and morphology complexity, etc. [3]. On the other side, unlike English langue the noticeable lack of Arabic language libraries that support NLP, does not encourage non-specialist researchers in the Arabic language to engage in research related to the Arabic language, because it will be complex and arduous for them.

Recently, Google AI Language researchers published a new pre-training language representations method called Bidirectional Encoder Representations from Transformers (BERT), It trained a language understanding model on a very large text corpus [4]. Such models have proven highly efficient at language understanding by achieving convincing results in most NLP tasks [5].

Also, at the start of this year 2020, the researchers published a new Arabic language model ARABERT based on BERT, and they evaluated this model on this filed Named Entity Recognition, Sentiment Analysis and Question Answering.

This is what made the research topic so interesting, to evaluate the efficiency and effectiveness of this model in automated Arabic text summarization using the extractive approach.

II. Related Works

There are several approaches for text summarization are proposed in research papers, like semantic -based, numerical-based, and hybrid approach. These approaches used for sentence scoring and selection [6].

A. Semantic-based Approach

Semantic technique models represent the conversation structure and consistency of texts using a Rhetorical Structure Theory (RST) [7]. RST is a theory that discovers the rhetorical relations between different text paragraphs and identifies the important ones, which used for Arabic summarization in [7]. This technique is separated into a few phases as follows. First, identify the conversation-units are related to each other based on predefined Arabic rhetorical relations. Then, building rhetorical relationships based on conversation phrases, and the

RS-trees are represents based on these relationships. Finally, the best RS-tree is chosen to generate the summary. This approach works fine with small and medium texts, but it suffered from large texts [7].

B. Numerical-based Approach

The Numerical-based approach is the most used and popular among other approaches, by assigning numerical scores to text sentences or words, then the summary generated based on high scores [9]. This approach can be archived via multiple methodologies like Word Frequency and Term Frequency-Inverse Document Frequency (TF-IDF).

- **Word Frequency methodology**: it is the most used methodology for sentence scoring, the sentence score is calculated by the sum of the frequencies with avoiding all stop words. The proposed solution in [10] generate news titles by combining Word-Frequency, sentence position and similarity measures methodologies.
- **Term Frequency-Inverse Document Frequency (TF-IDF)**: is a numerical methodology represents how the word is important to document in a collection of documents (corps) [9]. TF-IDF is better than the Word-Frequency in how the weight of a word in a document. TF-IDF is used frequently in auto-summarization systems [11-14]. In [11] proposed a solution that depend on inner product between the (TF) in a sentence and the (DF) for each extracted noun with (72% for recall and 62% for precision). In proposed solution [12] used TF-IDF on clustered word roots and achieved a good accuracy (with 78.7% for recall and 75.7% for precision).
- **Clustering**: is an algorithm used to extract the important sentences and remove redundancy from text to produce an efficient summary. In [15] clustered the similar sentences in one group to avoid it from the selection process, then select sentences with a high score from each cluster to reduce redundancy, and avoiding select sentences from the same cluster at the same time. Also, in [16] they extract and assign scores to the most important and unique terms in the document to reduce the redundancy and increase the relevance.
- **Machine Learning:** there are many techniques in this filed that can be used for sentences scoring like seq2seq, decoder encoder techniques and etc. In [17] they proposed a semi-supervised technique for extracting a summary and using it for training Seq2Seq models, and they found that RNNs and other Seq2Seq models are powerful and beat other summarization approaches especially in e-Commerce texts.

C. Hybrid Systems

In this field, researchers worked hard to combine text summarization approaches and techniques and take their advantages to improve results accuracy. In [18] the researchers proposed an extractive methodology that combines semantic and statistical features to evaluates every sentence in the text. Moreover, they used supervised machine learning and score-based techniques to generate an enhanced summary. On other hand, in [19] they proposed a solution for the English language by tokenizing the text into clean sentences, then the tokenized sentences passed to the BERT model to generates the embeddings, then using K-Means the embeddings has clustered. Finally, the summary selected based on the embedded sentences that were closest to each cluster centroid.

III. *Material and Methods*

The main goal of this research is proposing and evaluating a single Arabic document summarization, by modifying and extending the solution in [19] to accept and work with the new AraBERT model and compare the result with other text summarization approaches.

A. Input Preparation

Before passing the input text to the AraBERT model, we did a simple text preprocessing operation by removing Taskeel and Tatweel then we defining the sentence boundary by breaking the text into sentences. Then we remove the very short sentences from prepared text because it not meaningful and affects negatively the final summary.

B. Sentence Embedding

BERT has been chosen because it has a better performance than other NLP algorithms for sentence embedding. AraBERT followed the original BERT pre-training objective, and it's used the Masked Language Modeling (MLM) to improve the pre-training operation by forcing the model to predict the whole word instead of getting indicates from parts of the word. Also, it used the Next Sentence Prediction (NSP) to help the model understand the relationship between sentences [5].

Using AraBERT pre-trained model, the cluster learning system (CLS) embedding layer has chosen, this layer generates a required matrix (N x E) for the clustering process, which N represents the number of sentences and E represents the dimensions of embeddings. But unfortunately, the results of the (CLS) layer often not necessarily create the best embedding representation for sentences. Also, AraBERT model in other layers in the network produced another matrix (N x W x E), which W represents the tokenized words. To avoid this problem, we take the average of embedding to create a 2D matrix (N x E). Finally, now we can get the best embedding representation for sentences.

C. Clustering embeddings

Following the solution in [19], due to the performance similarity between K-Means and Gaussian Mixture Models, the K-Means model was selected for the clustering the generated embeddings matrix from the AraBERT model. Finally, from clustering results, the sentences closest to each cluster centroid were selected to be a proposed summary.

IV. EXPERIMENTS AND RESULTS

A. Dataset

Evaluating the auto summarization text is very difficult because there is no standard or specific summary for a given text. Also, the lack of an Arabic extractive summarization dataset made the evaluation process more difficult and subjective. So that we created a reference contains multiple articles in deferent topics with its summary, each summary generated by persons specializing in the Arabic language.

B. Evaluation Measure

There are two ways to evaluate the generated summary, first one is human-based, at this way the human extract the most important sentences form text then compares it with the generated summary. But it is an Impractical way because it is subjective and needs a lot of time and effort. On the other hand, the automatic-based evaluation is faster and depends on clear evaluation measures like (recall, precision, and f-score). ROUGE is the most popular automated measure used in text summarization, which is stands of Recall-Oriented Understudy for Gusting Evaluation [20]. It's used to evaluate the quality of the generated summaries by comparing it to its references.

C. Evaluation of Proposed Solution

The proposed solution was evaluated based on the two ways, first step we evaluated each article in our dataset separately, then take the average of them to be a model evaluation. On other hand, we selected a random article from our dataset and evaluated it manually by comparing them with human write summaries and take the average source to be a model evaluation. The "Fig. 1" and "Fig. 2" shows the ROUGE-1 and ROUGE-2 metrices (recall, precision, and F-measure). Moreover, we put the human evaluation at the end of table.

Approach	**ROUGE-1**			**Human eval.**
	recall	***precision***	***F-measure***	
AraBERT	0.39	0.90	0.54	0.52

Fig. 1. Proposed solution evaluation results using ROUGE-1

Approach	**ROUGE-2**			**Human eval.**
	recall	***precision***	***F-measure***	
AraBERT	0.37	0.83	0.51	0.52

Fig. 2. Proposed solution evaluation results using ROUGE-2

D. Comparison with Related Approaches

In this section, the proposed results were compared with other extractive summarization approaches using the same Arabic summarization dataset, and we found there is a difference from the evaluations recorded by the other researchers, and this shows that the evaluation process of summarizing the texts is difficult and very subjective. In order to better understand the results, we presented some random results to a specialist in the Arabic language, and the competition determined between Seq2Seq and the proposed approach. In this case, Seq2Seq approach extracted the paragraphs as all, and skipped other ones, so that generated summary was long. On another hand, the proposed approach generated a good summary by extracting the most important sentences from paragraphs and generated a coherent and meaningful summary.

However, according to our experiments, the proposed solution has a weakness. It depends on determining the sentence boundaries. So, if the text is not written well with careful punctuation used, this will negatively affect the results. Also, in the long texts, we have noticed there is a weakness in covering all important information, because the proposed solution selects the most important sentence in each group resulting from the clustering task. Finally, to make the generated summary more readable and understandable we should solve the issue of the linguistic expression by finding all these expressions and replace it with the referring entity. "Fig. 3" and "Fig. 4" shows the comparison result between them.

Approach	**ROUGE-1**		
	recall	**precision**	**F-measure**
Proposed solution	0.39	0.90	0.54
Word frequency	0.18	0.44	0.28
TF-IDF	0.26	0.37	0.30
Machin learning			
Seq2Seq	0.40	0.48	0.44
Decoder-Encoder	0.26	0.70	0.38

Fig. 3. Comparison between different approaches results using ROUGE-1

Approach	**ROUGE-2**		
	recall	**precision**	**F-measure**
Proposed solution	0.37	0.83	0.51
Word frequency	0.10	0.25	0.14
TF-IDF	0.16	0.23	0.19
Machin learning			
Seq2Seq	0.28	0.34	0.30
Decoder-Encoder	0.22	0.60	0.32

Fig. 4. Comparison between different approaches results using ROUGE-2

To take a look at the proposed solution efficiency, let take this example from online Arabic news website (Aljazeera) at this link. This article talked about the latest news for world leaders who were infected with the Coronavirus.

Proposed solution output:

"أعلن اليوم الخميس في ماليزيا عن دخول ملك البلاد في حجر صحي بعد إصابة 7 عاملين في القصر بفيروس كورونا، ليكون بذلك أحدث زعماء العالم التحاقا بقائمة القادة الذين تحوم حولهم شبهة الإصابة بهذا الفيروس. ويوم أمس الأربعاء، أعلن مقر إقامة ولي العهد البريطاني الأمير تشارلز إصابة الأمير بفيروس كورونا. وخضع الرئيس الأميركي دونالد ترامب لفحص تأكد بعده أنه غير مصاب، وذلك بعد لقائه بوفد برازيلي أحد أفراده مصاب بالفيروس. وفي أستراليا، نقل وزير الداخلية بيتر دوتون إلى المستشفى بعدما ثبتت إصابته بفيروس كورونا، بينما أعلن في إندونيسيا أن وزير النقل بودي كاريا سومادي نقل إلى المستشفى عقب إصابته بالفيروس.

وقالت صحيفة لوموند الفرنسية إن بوركينا فاسو هي البلد الأكثر تضررا بوباء كورونا (كوفيد-19) حتى الآن في غربي أفريقيا، حيث توفيت النائبة الثانية لرئيس البرلمان وأصيب خمسة وزراء، إلى جانب الحديث عن إصابة كل من السفير الإيطالي والأميركي، مما أثار غضبا على شبكات التواصل الاجتماعي بسبب ما اعتبر "تراخي" الحكومة في إدارة الوباء. ولم يتأخر وزير الشؤون الخارجية ألفا باري كثيرا بعدهم، حيث قال بعد يومين فقط من نفي الإشاعة رسميا إنه مصاب بالفيروس لقد تحققت الشائعات، تلقيت للتو اختبارا إيجابيا لكوفيد-19."

V. *Future Works*

This approach can be enhanced. So, we are going to work more to improve determining sentence boundary accuracy in the Arabic language. Also, putting a new layer for determining the sufficient number of sentences to generate the summary to solve very long text summarization issue. The other thing, the coreference resolution in the Arabic language is a hot topic to research, it can help us to reduce the ambiguity in generated summary by replacing refer expression with its name entity. Finally, the generated summary represents the highlight sentences in the text, so that we are going to use the reinforcement learning to convert it to an abstractive summary.

VI. Conclusion

The increasing sources of knowledge and the textual contents on the Internet need a powerful tool to generate a coherent and meaningful summary by extracting hotkey sentences and remove the redundancy from it, to try to cover the important information and helps the readers to find the main points of the text. This paper proposed and evaluated a single document extractive Arabic text summarization by combining NLU using (AraBERT) model and clustering technique. The proposed solution was evaluated by ROUGE-2 and achieved F-measure score 0.51, and to let this evaluation is more realistic, it is evaluated manually by a specialist in the Arabic language and achieved F-measure score of 0.52. But it has some weakness like it highly depends on sentence boundary, the coverage accuracy is decreased when the text is too long and the extracted sentences sometimes contain a linguistic expression that lets the summary has misunderstood. Anyway, we expected more from this approach, because it generated a coherent and meaningful summary.

References

[1] Radev, D., Hovy, D., McKeown, K. "Introduction to the special issue on summarization". Comput. Linguist 28 (4). (2002)

[2] Andrew Turpin, Yohannes Tsegay, David Hawking, and Hugh E Williams." Fast generation of result snippets in web search". In Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval. (2007)

[3] A. Abu-Errub, A. Odeh, Q. Shambour and O. Hassan, "Arabic Roots Extraction Using Morphological Analysis", in JCSI International Journal of Computer Science Issues, Vol. 11, Issue 2, No 1, March 2014.

[4] Jacob Devlin, Ming-Wei Chang, Kenton Lee, Kristina Toutanova "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding", 24 May 2019

[5] Wissam Antoun*, Fady Baly*, Hazem Hajj "AraBERT: Transformer-based Model for Arabic Language Understanding", 30 Mar 2020

[6] Lamees Al Qassem, Di Wang, Zaid Al Mahmoud, Hassan Barada, Ahmad Al-Rubaie, Nawaf I. Al moosa Automatic Arabic Summarization: A survey of methodologies and systems, 3rd International Conference on Arabic Computational Linguistics, ACLing 2017, 5-6 November

[7] W. Al-Sanie, A. Touir and H. Mathkour. "Towards an infrastructure for Arabic text summarization using rhetorical structure theory". M.Sc. Thesis, King Saud University, Riyadh, Saudi Arabia. (2005)

[8] Elghazaly, T., Ibrahim, A., "Arabic text summarization using rhetoricalstructure theory". International Conference on INFOrmatics and Systems. (2012)

[9] Ferreira, Rafael, et al. "Assessing sentence scoring techniques for extractive text summarization". (2013)

[10] Alotaiby, Fahad, Salah Foda, and Ibrahim Alkharashi. "New approaches to automatic headline generation for Arabic documents." Journal of Engineering and Computer Innovations 3. (2012)

[11] Al-Radaideh, Q., and Mohammad Afif. "Arabic text summarization using aggregate similarity." The international Arab conference on information Technology. (2011)

[12] Haboush, Ahmad, et al. "Arabic text summarization model using clustering techniques." World of Computer Science and Information Technology Journal. (2012)

[13] El-Ghannam, Fatma, and Tarek El-Shishtawy. "Multi-topic multi-document summarizer". (2014)

[14] Fejer, Hamzah Noori, and Nazlia Omar. "Automatic Arabic text summarization using clustering and keyphrase extraction." Information Technology and Multimedia (ICIMU), 2014 International Conference on. IEEE. 293-298. (2014)

[15] El-Haj, Mahmoud, Udo Kruschwitz, and Chris Fox. "Exploring clustering for multi-document Arabic summarisation." Asia Information Retrieval Symposium. Springer Berlin Heidelberg. (2011)

[16] Oufaida, Houda, Omar Nouali, and Philippe Blache. "Minimum redundancy and maximum relevance for single and multi-document Arabic text summarization." Journal of King Saud University-Computer and Information Sciences 26 (4): 450-41. (2014)

[17] Chandra Khatri, Gyanit Singh and Nish Parikh "Abstractive and Extractive Text Summarization using Document Context Vector and Recurrent Neural Networks" arXiv:1807.08000. (2018)

[18] Aziz Qaroush, Ibrahim Abu Farha, Wasel Ghanem, Mahdi Washaha and Eman Maali. ("An efficient single document Arabic text summarization using a combination of statistical and semantic features" Journal of King Saud University – Computer and Information Sciences. (2019)

[19] Derek Miller. "Leveraging BERT for Extractive Text Summarization on Lectures" arXiv:1906.04165. (2019)

[20] Lin C.Y.. Rouge: A package for automatic evaluation of summaries. In: Text Summarization Branches Out, http://www.aclweb.org/anthology/W04-1013. (2004)