



BERT Fine-tuning For Arabic Text Summarization

Khalid N. Elmadani*, Mukhtar Elgezouli*, Anas Showk

*These authors contributed equally

Abstract

Fine-tuning a pretrained BERT model is the state of the art method for extractive/abstractive text summarization, in this poster we showcase how this fine-tuning method can be applied to the Arabic language to both construct the first documented model for abstractive Arabic text summarization and show its performance in Arabic extractive summarization. Our model works with multilingual BERT (as Arabic language does not have a pretrained BERT of its own). We show its performance in English corpus first before applying it to Arabic corpora in both extractive and abstractive tasks.

Introduction

Arabic, one of six official languages of the United Nations, is the mother tongue of 300 million people, and the official language for 26 countries; nine of those are in Africa. Hence Arabic had a huge influence in mother Africa forming the culture and religious values in West Africa, consequently it's safe to say Arabic is "the Latin of Africa".

English is the golden standard for text summarization, strongly because of the vast number of well proposed benchmark data-sets containing a huge capacity of summarized articles both in extractive and abstractive schemes like CNN/Daily-Mail news highlights data-set (Hermann et al., 2015) (contains 287K news articles and associated highlights), this type of rich corpus is what Arabic language lacks in automatic text summarization. The lack of Arabic benchmark corpora makes evaluation for Arabic summarization more difficult. Without unified benchmark corpus, the results reported from existing model can only be a hint for overall performance comparison (AlQassem et al., 2017). But recently there is a turn out to use some corpora like KALIMAT: a Multipurpose Arabic Corpus (containing 20,291 articles with their extractive summaries).

Methodology

We used the pretrained BERT (Devlin et al. (2018)) for both abstractive and extractive summarization. The encoder (BERTSUM) (LiuLapata, 2019) is pretrained BERT, For abstractive summarization task the decoder is 6-layered Transformers (Vaswani et al., 2017) initialized randomly. This mismatching between encoder and decoder -the encoder was pretrained while decoder is not- may lead to unstable training, so Liu & Lapata (2019) proposed a new fine-tuning schedule which adopts different optimizers for the encoder and the decoder (BERTSUMABS). And, for extractive summarization task a sigmoid classifier was inserted on top of each sentence representation in the encoder indicating whether the sentence should be included in the summary (BERTSUMEXT).

This method of using pretrained BERT is perfect for our condition, because the pretrained model will compensate for the rela-

tively small data-set we are using. But how could this model be applicable for Arabic Language since BERT was trained on English documents? The answer is Multilingual BERT (M-BERT). It's similar to the normal BERT but trained on 104 languages. We trained one time using BERT and M-BERT another time on the CNN data-set for 45,000 steps to compare the impact of using M-BERT instead of BERT, sense M-BERT supports Arabic.

Finally, we included a non-pretrained Transformer baseline for both extractive and abstractive tasks, in order to measure the effect of using pretrained M-BERT. Both TRANSFORMERABS and TRANSFORMEREXT encoders are 6-layered transformers, the rest of their architecture is the same as BERTSUMABS and BERTSUMEXT respectively.

Model	R1	R2	RL
BERT	30.45	11.62	28.02
M-BERT	30.35	11.33	25.3

Table 1: ROUGE F1 results on the CNN test set

Model	R1	R2	RL
BERTSUMEXT	42.02	24.59	41.99
TRANSFORMEREXT	28.75	14.80	28.72
BERTSUMABS	12.21	4.36	12.19
TRANSFORMERABS	6.93	1.78	6.88

Table 2: ROUGE F1 results on KALIMAT test set

Results

Table 1 demonstrates the first step towards Arabic text summarization; switching from monolingual BERT to multilingual BERT. The results show very similar performance as compared to BERT and M-BERT.

Table 2 presents our results on KALIMAT data-set. We conclude that pre-trained M-BERT leads to huge improvements in performance for relatively small data-sets in both extractive and abstractive summarization. It also reveals that extractive models would have higher performance for extractive data-sets than their corresponding abstractive ones.

Conclusion

In this poster, we showed how multilingual BERT could be applied to Arabic text summarization and how effective it could be in low resource situations. Research in Arabic NLP is still in its infancy compared to English; abstractive text summarization was not attempted before (at the time of this submission) so there is no metrics output that we can evaluate against.