

Abstractive Text Summarization using Seq2seq Model

Keerthana S.
ME CSE
PSG College of Technology
Coimbatore, Tamilnadu

Venkatesan R., PhD
Professor
PSG College of Technology
Coimbatore, Tamilnadu

ABSTRACT

Knowledge is power and information is liberating. As the quote says, in today's world the information is available in abundance and a lot of new possibilities can be explored from them. Text summarization is one of the main applications of natural language processing. Text summarization is one of the widely used methods to process the text corpus and obtain a precise text that captures the entire context and preserves the important information conveyed through the text. This paper presents an approach of abstractive text summarization using the seq2seq model. The proposed methodology aims at enhancing the efficiency of the summary generated with the help of the data augmentation technique. The summary comprises new words and sentences thereby improving the quality of it. For evaluating the quality of summarization bilingual evaluation understudy (BLEU) score is used.

General Terms

Abstractive summarization, seq2seq, attention mechanism.

Keywords

Synonym replacement, LSTM, Wordnet.

1. INTRODUCTION

In today's world, the amount of data that is processed every day is very large. The task of processing such data is very challenging. Text summarization provides a solution for summarizing these large documents to obtain short and precise summaries [1]. There are two approaches to text summarization: extractive and abstractive [2]. Extractive summarization produces summaries by extracting keywords, phrases, and sentences from the source document and combining them. The summaries generated preserve the actual context of the source article [3, 4]. Abstractive summarization, on the other hand, produces a summary that mimics the human written form. The summaries ought to contain new words and sentences that are not present in the source text [5, 6]. In this way, verbally innovative sentences can be added to the summary which enhances the quality of the summary. Compared to the extractive approach, the abstractive approach is complicated. Deep neural network models have been proving to perform better for the abstractive approach of summarization. The work is primarily focused on the abstractive summarization using the seq2seq model.

Seq2seq models have been used for many tasks in natural language processing such as machine translation, speech recognition, video captioning, etc. The seq2seq model consists of two main parts the encoder and the decoder [9]. The main function of the encoder is to encode the source text to the context vector that preserves the information provided in the source text. The function of the decoder is to generate the target word for each time step according to the context vector produced by the encoder. But the baseline models had many issues like Out of vocabulary deficiency, disfluency, repetitive words in the summary. In order to overcome such problems,

they came up with the attention mechanism. The attention mechanism produces an attention vector which helps the decoder by showing which parts of the context vector must be more focused to produce a summary that preserves the context of the source article [10]. The decoder is trained on teacher forcing methodology. It is forced to produce a word that is similar or the same as the target word. Hence with the help of data augmentation, the words of the articles that are to be trained on are replaced with their respective synonyms. In this way, words are changed and the context vector for the article is obtained based on the changed sentence and the decoder is forced to generate words that are similar to it. So new words can be introduced into the summaries generated and more fluent, grammatical sentences can be obtained from the model after the process of training.

2. RELATED WORK

In 2010, Foaad khasmod et al. [7] proposed a style transformation technique that demonstrated a great impact on the text transformation. Followed by pre-processing, every word in the text is provided with a part-of-speech tag. Every sentence is then divided into sentence fragments in such a way that they do not overlap. Each fragment will be examined for replacement with the help of Wordnet. Wordnet is a lexical database that comprises of synsets. The synsets are one or more groups of synonyms for a particular lemma. So every word's synonym is looked upon the synsets. Each synonym is ranked based on its hit rate. The word with maximum score is chosen and replaced with the text.

In 2016, Ramesh Nallapati et al. [5] proposed several models for abstractive summarization. The baseline model consists of an encoder and a decoder. It is constructed with the help of Gated Recurrent Unit-RNN. The encoder is bidirectional and the decoder is unidirectional. The major change to this model is the decoder vocabulary. It is restricted to the words in the source document for that particular batch. In this way, the size of the soft-max layer of the decoder is reduced. The model captures the key concepts and entities with the help of additional lookup based embedding matrices. It captures the linguistic features efficiently. To handle the out of vocabulary words, generator pointer mechanism is used. The decoder checks whether the word is present in the training data. If so, it is considered for processing else it is pointed in the source document and later on considered for summary generation. All the above-mentioned changes are adapted to the basic model to produce efficient summaries.

In 2018, Yong Zhang et al. [8] proposed a framework for key phrase generation using the seq2seq model. The baseline architecture comprises bidirectional Gated Recurrent Unit and Unidirectional Gated Recurrent Unit as encoder and decoder respectively. The main objective of this framework is to deal with Out of vocabulary words and the repetition that occurs in the summary. Out of Vocabulary (OOV) problem is overcome by the copy mechanism. All words are split into two parts-

fixed vocabulary that contains the most frequently occurring words and OOV vocabulary. While generating a summary, the decoder uses this mechanism as a soft switch to choose a word from the fixed vocabulary or by copying word directly from the source based on probability distribution. Repetitions are handled by coverage mechanism. It can be viewed as a memory vector and hence the phrases generated at each time step will have less similarity thereby avoiding repetition.

In 2019, Jianwei Niu et al. [9] proposed a framework of abstractive summarization that is based on the seq2seq model. A novel attention mechanism namely, the Sun attention mechanism was introduced to learn the context vector efficiently. In the basic seq2seq model, the encoder consists of the bi-directional LSTM and the decoder consists of LSTM. The conventional attention mechanism only considers the encoder outputs and current hidden state of the decoder. It ignores the decoder input. The proposed Sun attention mechanism considers the encoder outputs and the decoder inputs to produce the distributions. Thereby summary generation takes place by focusing on both context information and the last word of the current state of the decoder.

3. PROPOSED METHODOLOGY

3.1 Pre-processing

Any natural language processing task related to text involves the pre-processing of the text. It involves the conversion of all words in lower case, expanding the contractions in the text, removal of punctuations, removal of special characters, and removal of stop words from the text. The processed text after all these steps is passed on for synonym replacement.

3.2 Synonym replacement

Synonym replacement is performed with the help of Wordnet. Initially, each word in a sentence is attached with a tag that specifies the part-of-speech of that particular word. Later on, words other than nouns are replaced with their respective synonyms with the help of synsets in the wordnet. Synonyms are thereby included in the text, which is passed on to the encoder for further processing.

3.3 Encoder

The encoder is made up of long short term memory units (LSTM). Among the variants of recurrent neural networks, LSTM overcomes the problem of vanishing gradients. Also, the LSTM's are efficient for understanding the context of sequences that are very long. In order to process the text properly, the encoder must produce a context vector that captures the entire context. Hence use of LSTM is necessary concerning the context. Stacked LSTM's are used in order to understand the sentences efficiently. Before encoding the sentence, the input text cannot be processed as such. It is converted into an integer format based on the term frequency in the sentence. Likewise, each sentence is converted into a sequence of integers based on their frequency and lookup dictionaries for each word along with their integer are built.

Only after building the vocabulary, the sentence that is converted into a sequence is passed into the encoder for obtaining the context vector.

3.4 Attention Layer

Encoder produces hidden states of each element in the input sequence. Alignment Scores are calculated between the previous decoder hidden state and each of the encoder's hidden states. The last encoder hidden state can be used as the first hidden state in the decoder. The alignment scores for each encoder hidden state are combined and represented in a single vector and subsequently softmaxed. The encoder hidden states and their respective alignment scores are multiplied to form the context vector. The context vector is concatenated with the previous decoder output and fed into the Decoder.

3.5 Decoder

The input to the decoder is the previous state output of the decoder, the hidden state of the previous time step, and the context vector obtained through the attention mechanism. The decoder calculates the target word probabilities based on these vectors at each time step. The greedy approach is followed for choosing the target word. Word with the highest probability is chosen and generated as the output of that time step. Likewise, an entire sequence is generated by the same mechanism.

4. EXPERIMENTS

4.1 Dataset

The dataset consists of 98K examples and contains Headlines, Short text of the complete Article. It consists of news articles from Hindu, Indian times and Guardian. Time period ranges from February to august 2017. The model mainly focuses on the summarization of news articles.

4.2 Evaluation metric

The model should perform very well on the training dataset and ideally have been generalized to perform well on the test dataset. Ideally, a separate validation dataset is used to help with model selection during training instead of the test set. Evaluation involves two steps: first generating an output sequence, and then repeating this process for many input examples and summarizing the skill of the model across multiple cases. The BLEU scores are calculated to get a quantitative idea of how well the model has performed. The Bilingual Evaluation Understudy Score or BLEU is a metric for evaluating a generated sentence to a reference sentence. This metric is based on n-gram score. The bleu score indicates how similar the candidate text is to the reference texts, with values closer to one representing more similar texts. The text summarization model built basically produces a summary and the evaluation is based on the similarity measure between the summary that already exists for the text taken as input and the summary that is produced as output.

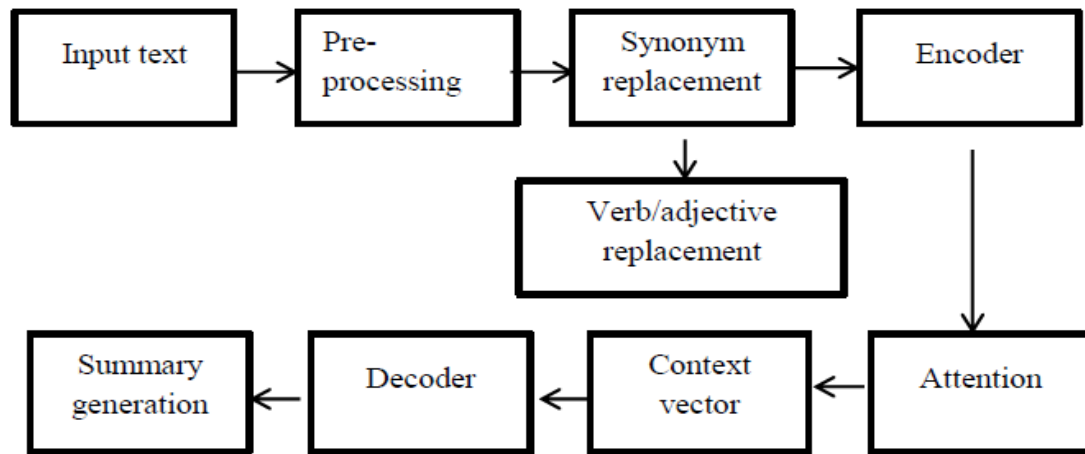


Fig 1: Proposed method

4.3 Implementation Details

The input and output vocabulary is the same. The words in the vocabulary are collected from the dataset after pre-processing. The embedding size of the word is 300. The batch size used is 128. The loss function used is sparse categorical cross entropy. The optimizer used for the process is rmsprop. The number of epochs ran for training the model is about 25.

4.4 Experimental Results

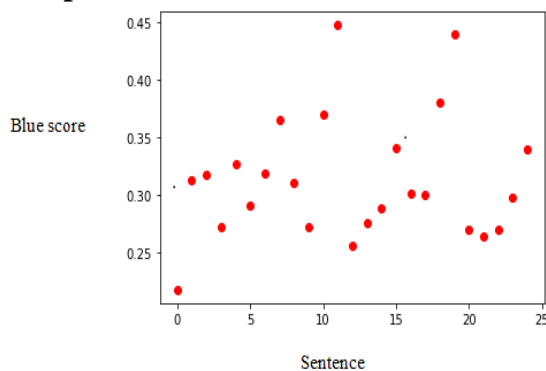


Fig 2: blue score comparison

BLEU score is computed by the average of unigram, bigram, trigram, and 4-gram precision. Also, the brevity penalty is considered that monitors the length of the reference sentence and the predicted sentence. Altogether, the bleu score signifies how much overlap is there between the reference text and the prediction. In the experiment, 25 sentences are validated and the score of each sentence is plotted. As per the plot, 75% of the scores ranged between 0.25 and 0.38. In this case, since news articles are experimented it is important for the primary information to be focused in the summary. The model proves effective in achieving the above mentioned factor. It signifies that the predictions are understandable and are not overlapping beyond a certain limit. This is due to the fact that there are new words in the prediction when compared to the reference. And these words do not alter the context of the source text.

5. CONCLUSION

Deep learning based approaches show promising results in solving abstractive summarization. The encoder-decoder model has been successfully applied along with attention

mechanism to obtain good results on summarizing the text. By calculating the attention vector, the model is manipulated accordingly to produce human-written like summaries. The future work is to improve the scalability and generalize large paragraphs to obtain summaries.

6. REFERENCES

- [1] Radev, D.R., Hovy, E. and McKeown, K., 2002. Introduction to the special issue on summarization. *Computational linguistics*, 28(4), pp.399-408.
- [2] Mani, I. and Maybury, M.T., 1999. *Advances in automatic text summarization* MIT Press.
- [3] Gambhir, M. and Gupta, V., 2017. Recent automatic text summarization techniques: a survey. *Artificial Intelligence Review*, 47(1), pp.1-66.
- [4] Verma, R. and Lee, D., 2017. Extractive summarization: Limits, compression, generalized model and heuristics. *Computación y Sistemas*, 21(4), pp.787-798.
- [5] Nallapati, R., Zhou, B., Gulcehre, C. and Xiang, B., 2016. Abstractive text summarization using sequence-to-sequence rnns and beyond. *arXiv preprint arXiv:1602.06023*.
- [6] Rush, A.M., Chopra, S. and Weston, J., 2015. A neural attention model for abstractive sentence summarization. *arXiv preprint arXiv:1509.00685*.
- [7] Khosmood, F. and Levinson, R., 2010, December. Automatic synonym and phrase replacement show promise for style transformation. In *2010 Ninth International Conference on Machine Learning and Applications* (pp. 958-961). IEEE.
- [8] Zhang, Y. and Xiao, W., 2018. Keyphrase generation based on deep seq2seq model. *IEEE Access*, 6, pp.46047-46057.
- [9] Niu, J., Sun, M., Rodrigues, J.J. and Liu, X., 2019, May. A Novel Attention Mechanism Considering Decoder Input for Abstractive Text Summarization. In *ICC 2019-2019 IEEE International Conference on Communications (ICC)* (pp. 1-7). IEEE.
- [10] Bahdanau, D., Cho, K. and Bengio, Y., 2014. Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:1409.04*.