WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Abstractive Text Summarization using Seq2seq Model

Dr.J.B.Jona, Mr.S.A.Gunasekaran, Aadhithian.L, Nithin.S.V, Pranaav.A.R

Department of Computer applicationCoimbatore Institute of Technology Coimbatore,India

ABSTRACT:

Knowledge is power and knowledge is liberating. As this quote points out, in today's world, information is abundant and there is a great deal of potential for innovation. Text summarizing is one of the main applications for processing natural language. Text summarization is one of the most widely used methods of copying text and obtaining accurate text that captures the essence and preserves the important information conveyed through the text. This paper introduces an abstract version of the text using the seq2seq model. The proposed approach aims to improve the efficiency of the summary produced with the help of a data augmentation strategy. The abbreviation combines new words and sentences thus enhancing its quality. To assess the quality of the summary of the bilingual test standard in understudy (BLEU) is used.

General Terms: Abstractive summarization, seq2seq, attentionmechanism.

Keywords: Synonym replacement, LSTM

Introduction:

In today's world, the amount of data that is processed daily is enormous. The task of processing such data is very challenging. Summary text provides a solution forsummarizing these large texts for short and accurate summaries. There are two ways of abbreviating text: extractive and abstractive. The output summary generates summaries by extracting keywords, phrases, and sentences from the source document and compiling it. The abbreviations produced maintain the true context of the source article [3, 4]. Invisible summaries, on the other hand, produce a summary that mimics human writing. Abbreviations should contain new words and phrases that are not in the source text [5, 6]. In this way, word sentences can be added to the summary which improves the quality of the summary. Compared to the extraction method, the abstractiveextraction method is more complex. Models of thedeep neural network seemed to work best in a concise approach. The work is mainly focused on abstract concepts using the seq2seq model.The Seq2seq modelshave been used for many functions in natural language processing such as machine translation, speech recognition, video captions, etc. The seq2seq model consists of two main components an encoder and a decoder. The primary function of the encoder is to encode the source text in the context vector whichstores the information provided in the source text. The function of the decoder is to generate a target name for each step of the time according to the context vector generated by the encoder. But the basic models had many problems such as Without the lack of words, confusion, repetitive words in a summary. Span generates an attention vector that assists the decoder by indicating which parts of the context vector should be most focused on producing a summary that retains the context of the source article [10]. The decoder is trained by the teacher's method of coercion. It is compulsory to produce the same or similar word to which it is intended. So with the help of data additions,the names of the articles to be trained are replaced with words of the same meaning. In this way, the words are changed and the total vector of the article is obtained based on the changed sentence and the decoder is compelled to produce similar words. New words can therefore be introduced into abridged sentences and smooth, grammatical sentences can be found in the model after the training process.

Following a preliminary consideration, each word in the text is given a mark for part of the speech. The whole sentence is then divided into pieces of the sentence in such a way that they do not collide. Each piece will be tested for replacement with the help of Wordnet. Wordnet is a lexical website that integrates synsets. Synsets are one or more identical words for a particular lemma. So each same word is looked at in the synsets. Each similar name is rated based on its beat rate. The word with the highest score is selected and replaced with the text. In 2016, Ramesh Nallapati et al. [5] suggested a few examples of abstract concepts. The basic model consists of an encoder and an encoder. Built with the help of the Gated Recurrent Unit-RNN. The encoder is redirected and the encoder is directed at each one. A major change in this modelis the decoder vocabulary. Limited to the text in the source text of that particular collection. In this way, the size of the soft-max layer of the decoder is reduced. The model captures the key concepts andentities 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 inthe source document and later on considered for summary generation. All the above-mentioned changesare 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 repetitionthat 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 choosea 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 haveless 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 placeby focusing on both context information and the last word of the current state of the decoder.

MODELS

This work primarily focuses on adding context as a first step to RNNs to find abstractive and extractive text summaries and compare them with various high- level techniques. To extract it we use labeled and lightly labeled data. To quote we use titles and subtitles to train models.

A. Nomenclature and Basic Model

Recurrent Neural Network is a type of neural network that is an extension of the NN Forward, which has at least one response connection, so that the startup flowsin the loop. Basically, information from previous observations and current observations is used to make predictions.

Theoretically, this framework should work but it is found that RNNs have coding mechanisms that find it difficult to map long sequences or where there are long-term dependencies. Gated Recurrent Units (GRUs) [8] and Long-Term Memory (LSTMs) [34] solve this problem by introducing gateways to the network to prevent the gradient extinction problem associated with long-term RNNs. In the present study, LSTMs were used to summarize the output and the invisible. In standard LSTM fixed lengths are transmitted as coded in a fixed dimension vector (v), and then separated by a output sequence of words. In summary, LSTM estimates the following:

$p(y1, y2, \ldots, yT' | x1, x2, \ldots, xT) = p(yt | v, y1, y2, \ldots, yt-1)$

With unambiguous summaries, the document can be served as input during training and summaries can be served as outputs. However, the RNN output can be trained using standard guarded setting by doing soft- max in a coded layer. Contextual Recurrent Neural Network.

Abstractive Contextual RNN (AC-RNN).

The RNN structure for displaying the contents of a document as described in Section 3 is approved as the first input and the sequence of the document in the encoder. The idea is that if the previously read- document-vector (vd) document is passed as input at the beginning of the text section, then the model not only changes quickly but also reads the abbreviations that accompany the document and not just the usual sequence. The basic premise is that if the reader knows the title of the document or the abstract of the book, then it provides a better understanding and higher definition of the text that the model is able to provide. Straight forward summaries accompanying documents. Therefore, the document-context-vector(vd) in the first step changes the encoder vector.

The LSTM decoder with the same format mentioned inencoding was used, however the t = 0 input encoder is a vector detected in the encoder [6, 37]. Unlike other coding methods, where the output time t can be any word from the vocabulary, the output from the document name information is considered at the time of the prediction, which makes the reflection faster.

Sutskever proposed a beam search to find the most likely sentence in a typewriter work. However, we do heuristics by using a trained model to re-standard all sentences (within the document) based on the ability to make a sentence become a summary sentence during suggestion. When the possibilities for any sentence are defined as coding opportunities the sentence is given a coded input. Therefore, the abstractive model is used for extraction during the description. We have adopted a view based on the advanced method of Automated Speech Recognition, in which RNN is used to re- evaluate potential outcomes from a n-gram-based language model. We do this to address a number of issues: (a) Avoid common problems and short releases in sequence of consecutive models, (b) Find grammatically correct sentences for eBay users to avoid negative customer information and knowledge about customer concerns. (c) Avoiding legal push- backs from the sellers

Extractive Contextual RNN (EC-RNN)

It contains only Encoder. The encoder used in EC- RNN is an encoding code used in AC-RNN with document-context-vector (vd) as input time t = 0 and embedding word representation transferred input to model. However, the code encoder output is used for dual split (sentence is a summary sentence or not)using softmax. Note that each sentence starts with a document- context-vector, so sentence breaks occurwhen viewing the context of a document and not just the sentence itself. In this way the same sentence can be divided into an abridged sentence

Total vocabulary size	768,298
Median documentlength	346 characters

Median number ofwords	54
Median sentence length	51 characters
Median Number of words in sentence	8

Table 2.1 Data Details

one document but not for others. Furthermore, given a document- context-vector (vd), it is the extra information provided by the sentence which differentiates it from the other sentences in the document, which is a major drawback of other state- of-the-art classification approaches wherein some sentences are always classified as a summary sentence.

Non-Contextual RNN Architectures

In the current setting, RNNs trained outside the document-context-vector are called non-contextual RNNs. Recently, a number of structures have been proposed in this regard [1, 7, 15, 21, 28, 32], however three main models are examined in the present study: Abstractive RNN (A-RNN). Abstractive RNN is a traditional sequence modeling model using LSTM proposed by [6, 37]. The model is exactly the same as AC-RNN without the content as input during t = 0.Input time t = 0 in A-RNN is a symbol.

<start>. Limited input and output sequences are used for reduction training or paving .Extractive RNN (E- RNN). RNN has been used for classification tasks and it generates state-of-the art results. RNN rendering is a non-contextual version of the proposed EC-RNN As mentioned earlier, embedding is pre-calculated using the Skip Gram With Negative Sampling (SGNS) method and is used as the corresponding input for each word in the coding layer. Elements are extracted using embedding function separator.

Convolutional RNN (CNN-RNN).

Convolution based LSTM has performed extremely well in text classification tasks. Furthermore, convolution attention-based encoder has been used for short summarization tasks. CNN-LSTM is used to classify the sentences with the same technique as E- RNN, however the difference is that CNN is used to extract sequences of higher- level phrase representations. As suggested by Zhou et al. [42] CNN- LSTM is able to capture both local features of phrases, global and temporal sentence semantics. CNNs with multiple filters, max-pooling and dropout are used to extract high-level phrase representations and then passed to LSTM for classification using Softmax.

DATASETS

Table 1 describes the distribution of a eBay description. Vocabulary size of our dataset is 768K words. Median document length is 346 and 54 words.

There are two types of data sets used in the current study. Recorded People: 20,000 items / documents and related information (titles, url, description) were provided to the public for summary. The task was to extract and rate the sentences in the descriptions provided by the eBay object url. 5k out of 20k items were used for testing (Gold Set) and 15,000 items were used to train different models.

An abbreviated scale for a slightly larger scale sentences from definitions of 100,000 objects were extracted and rated based on the context of the document. Part

3.4 provides information on how to measure a sentence in the order of importance given to the document and its context details. Several techniques suggested by Shen et al., Lin et al. and [2, 13] based on question, theme similarity, title signature and Latent Semantic analysis. [29] expanded articles on Wikipedia and found the best results of DUC's summary work.

Similar to the methods mentioned above, we find almost identical descriptions of eBay object descriptions. Following a review from eBay reviewers, it is found that summaries generated using text-based content are high quality and can be used to train models. Considering the quality of these summaries, eBay introduced this feature to mobile applications and websites. In A / B testing, it is found that displaying summaries using this method has a higher cost compared to non-displaying snapshots. Models trained using a low-monitoring method will be tested on the Gold test set to determine the relevance of this process.

Data Generation for Classification task

EC-RNN, C-RNN, CNN-RNN and other class summary techniques require data labeled training. In classification tasks, sentences with a restricted list are labeled as an abridged paragraph while sentences that get high marks in the document content content are considered positive sentences. Restricted listing words and phrases that do not contain object / document level information and are commonly used on eBay such as "return", "post", "5 star rating", etc. We have acquired 700 words using personal care and statistical analysis of eBay object descriptions. In each description of 100,000 objects, the sentences are marked positive or negative. Sentences that do not have high marks in the content of the document content or that contain the names of the restricted lists are omitted unless marked. As, the context of the document context is a new approach and will be tested without using A / B testingin production, the data is marked with high accuracy. This activity is a step towards assessing the quality of sentences based on the content of the document.

ARCHITECTURE DETAILS AND EXPERIMENTATION

Training Details and Model Architectures Abstractive Context RNN (AC-RNN) and Abstractive RNN (A-RNN) were trained using deep LSTM with 4 layers (as described in Sutskever et al. [37]) with 1000 cells and 300 dimension word embeddings. Since, we wanted to find the relative difference after adding the context in RNNs for summarization task, we kept the same parameters for both the models. Parameters settings and model details which worked the best in our case are mentioned in Table 2.

C. Table 3.1 –Abstractive Approaches like AC-RNN and A-RNN

Parameter	**Value**
Input Description ength	**50 words**
Output SummaryLength	**15 words**
Optimization Method	**Stochastic GradientDescent with momentum**
Learning Rate	**0.1 ;reduced to halfafter every third epoch**
Batch Size	**128**
LSTM Parameter	**Uniform Distributionfrom [-0.1,0.1]**

D. Table3.2 –Extractive Approaches like AC-RNN and A- RNN

Parameter	**Value**
Maximumsentencelength	**15 words**
Optimization Method	**Adam**
Learning Rate	**0.1**
Batch Size	**256**
LSTM Parameter	**Uniform Distributionfrom [-0.1,0.1]**

Table 3.3-Parameter setting for Convolutional RNN

Parameter	**Value**

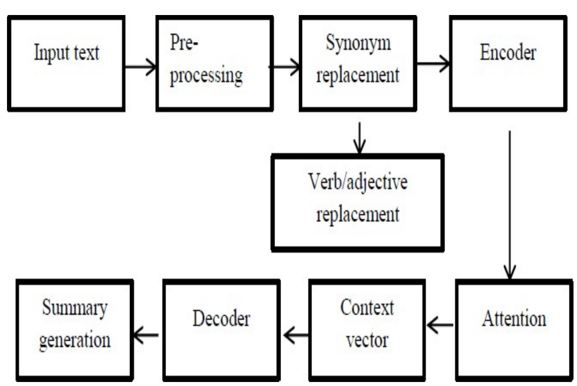
Dropout keepprobability	**0.5**
Maximum sentencelength	**15 words**
Learning Rate	**0.01**
Filter Size for Convolution	**4**
Batch Size	**128**
CNN and LSTMParameter	**Random Normalcentered at 0withstandard deviation 0.1**
Max Pool Size	**4**

Extractive Context RNN (EC-RNN) and Extractive RNN (E-RNN) were trained using two LSTM layers with 300 cells and 300-dimension embedding. Since, we wanted to find the related differences after adding context to the RNNs for the summary function, we kept the same parameters for both models. The parameter settings and details of the model that worked best for us are mentioned in the table above.

Convolutional RNN (CNN-RNN) contains two neural networks. CNN introduces high quality sentences and then LSTMs for temporary status and text sequence. We used Convolution for a single layer with a filter size of 4 and LSTM for Single layer with 300 cells and 300-dimensional word embedding. The parameter settings and details of the model that worked best for us are mentioned in Table

EVALUATION METRICS

The model should perform very well in the training database and be standardized in order to perform well in the test database. Ideally, a separate set of validation data is used to assist in choosing a model during training instead of a test set. Testing involves two steps: first producing the output sequence, and then repeating the process with multiple examples to include and summarize the model capability in all multiple scenarios. BLEU scores are calculated to get the maximum idea of how well the model works. The Bilingual Evaluation Understudy Score or BLEU is a matrix for evaluating a generated sentence to become a reference sentence. This metric is based on the n-gram result. The bleu effect shows how the candidate's text is similar to the reference text, the values closest to each other representing the same text. A text-based summary summary model generates a summary and the test is based on the ratio of the similarity between an existing summary of the text taken as inserted and the summary produced as output.



A. Implementation Details:

The input and output vocabulary are the same. Vocabulary words collected form a data set after pre- processing. The word embedding size is 300. The size of the collection used says 128. The loss function used is the entropy of different categories. The optimizer used in the process is rmsprop. The number of model training sessions is approximately 25.

B. Experimental Results

BLEU score is calculated by unigram, bigram, trigram, and 4-gram accuracy. Also, the abbreviated penalty is considered to monitor the length of the reference sentence and the predicted sentence. Overall, the bleu score indicates how much overlap exists between the reference text and the prediction. In the test, 25 sentences are verified and points for each sentence are constructed. As of structure, 75% of schools are between 0.25 and

0.38. In this case, as news articles are explored it is important that basic information is

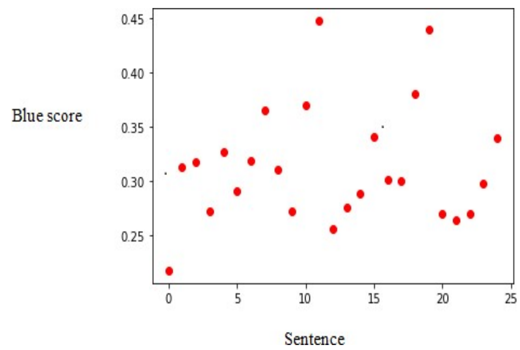


Figure 4.1 BLEU Score

focused on the summary. The model seems to be successful in achieving the aforementioned feature. It means that the predictions are understandable and do not go beyond a certain limit. This is because there are new words in the forecast compared to the reference. And these words do not change the context of the source text.

CONCLUSION:

In-depth study-based approaches show promising results in solving incomprehensible summaries. The encoder-decoder model has been used successfully with the attention-grabbing method to get the best results in abbreviated text. By calculating the attention vector, the model is effectively used to produce human-written abstracts. The work of the future is to develop scales and produce large sections to obtain summaries.

REFEERENCES:

[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 inautomatic text summarization MIT Press.

[3] Gambhir, M. and Gupta, V., 2017. Recent automatic text summarization techniques: asurvey. 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 preprintarXiv:1602.06023.

[6] Rush, A.M., Chopra, S. and Weston, J., 2015. Aneural 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 InternationalConference 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.