Proceedings of IYSC, (2021), vol. 10

Journal homepage: http://journals.sdu.edu.kz/index.php/iysw

COMPARATIVE ANALYSIS OF THE THREE STATE-OF-THE-ART TRANSFORMER-BASED SEQ2SEQ ABSTRACTIVE SUMMARIZATION MODELS

Ospan Smagul

Department of Computer Science, Suleyman Demirel University, Kaskelen

Email: 201107022@stu.sdu.edu.kz

Abstract

Today everyone can clearly see that amount of information on the internet growth rapidly. Because of it, people meet with huge problem - to process and get the most important parts from them. Thus, there is become a need on its clear and cost-effective summarization. Main goal of text summarization is to generate concise and accurate summary from input documents. In recent years, text summarization become one of the most topical subject in tech sphere. Big Tech companies like Facebook, Google, and Microsoft understood the importance of automatic summarization technologies, and not long ago published their results called BART, PEGASUS and ProphetNet respectively. This works showed the best results on various datasets with different sizes. The main idea behind this paper is to compare and analyze those models in terms of speed, accuracy, accessibility and other characteristics. In the next work, the results of the research are planned to be applied on Russian and Kazakh language datasets.

Keywords: summarization, BART, PEGASUS, ProphetNet, dataset.

Comparative analysis of the three state-of-the-art transformer-based seq2seq abstractive summarization models

Every day the amount of information on the internet is only growing and it becomes hard for human to process such a huge amount of data. Solution for this kind of problem is automatic text summarization, which can help to deal with such difficulties and allow identifying key ideas from large text easily.

In general, there are two approaches of automatic text summarization in NLP: extractive and abstractive. In extractive method, a subset of words that the most important in terms of full text pull from them, combine and form a summary. On the other hand, abstractive summarization approach more like a human did summary. Its implementation is based on advanced deep learning techniques, which applied to rephrase and shorten initial document, like human do. It is obvious that, in order to develop this kind of summarizer may take a lot of effort, as they would need the generation of natural language. This is main reason, why the abstractive summarization is more correct and difficult to realize than the extractive one. Recently, three Big Tech companies presented their solution of this kind of summarization algorithms: BART (Facebook), PEGASUS (Google) and ProphetNet (Microsoft).

The paper presents a detail survey of those three summarization techniques and advantages and limitation of each method. Description of each method is defined in section 1. Results of the experiment is discussed in section 2. Related works are discussed in section 3. Finally, section 4 concludes the paper.

I. Description of methods

BART is a denoising autoencoder, which is trained by corrupting text with a random noising function, and learning the model to reconstruct the original text. Actually, BART is based on BERT extractive text summarization algorithm and GPT summarizer, which is used for generation. As shown on Figure 1, the document is bidirectionally encoded, and then tokens are autoregressively predicted. It gives new state-of-the-art results on different tasks like summarization, abstractive dialogue or question answering [1].

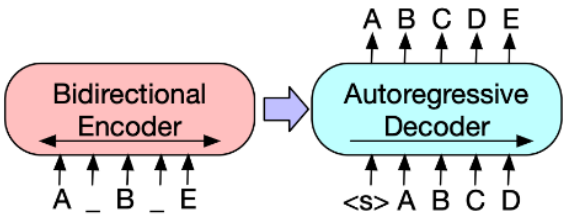


Figure 1. A scheme about how BART works

PEGASUS hides important sentences from an input document and similar to extractive summary, generate output sequence out of the remaining sentences. As shown on Figure 1, architecture of PEGASUS is based on Transformer-based encoder-decoder. It shows state-of-the-art results on all 12 datasets, which were used in official paper. Moreover, it achieves good results with only 1000 examples by showing state-of-the-art results on 6 datasets [2].

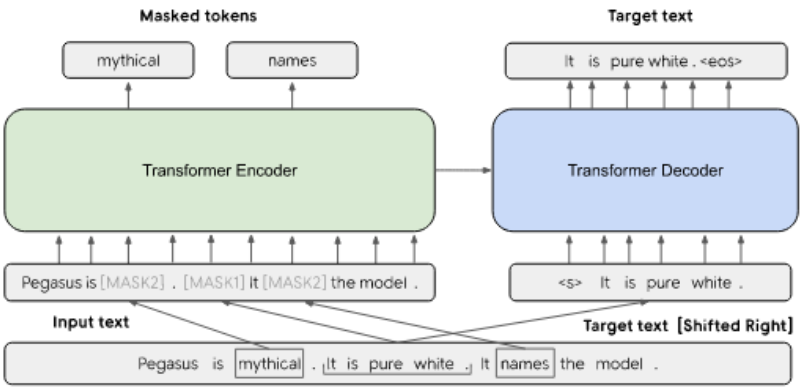


Figure 2. Standard Transformer encoder-decoder is the base architecture of PEGASUS

ProphetNet predicts the next n tokens at the same time based on previous context tokens at every time step. As shown in Figure 3, some token spans of the original text are masked out as the encoder input, and the model learns to recover those masked tokens. It achieves state-of-the-art results on CNN/DailyMail, Gigaword and SQuAD datasets compared to other models, which use the same scale pre-training corpus (16 GB and 160 GB) [3].

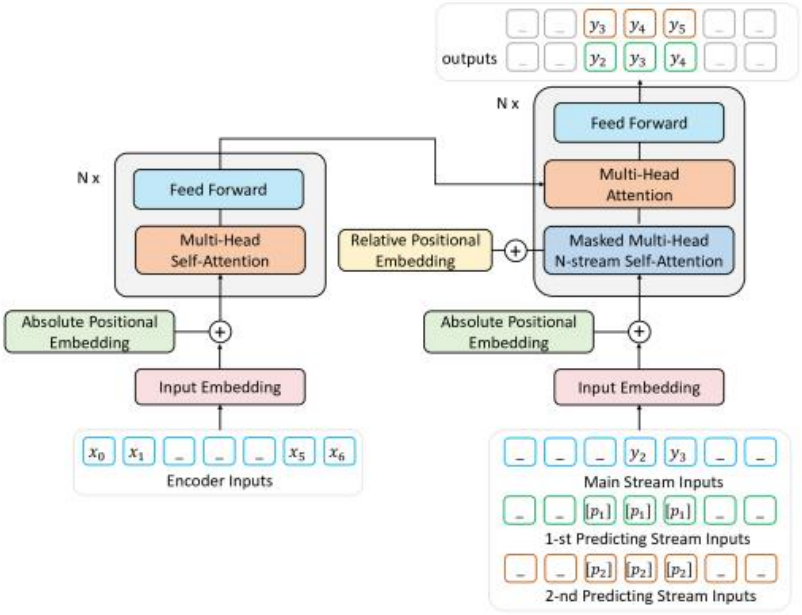


Figure 3. The architecture of ProphetNet

II. Results of the experiment

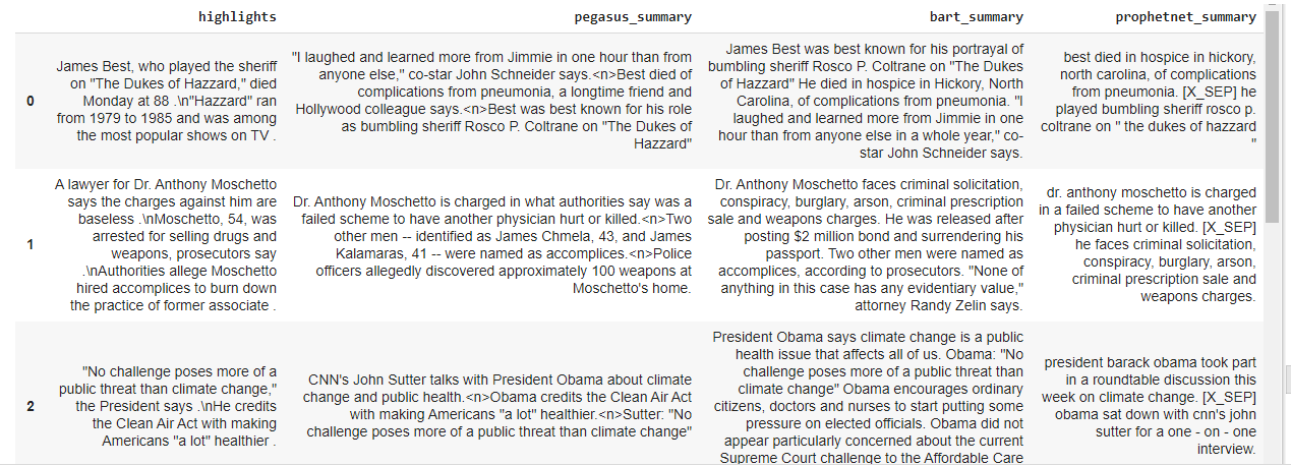
All experiments were done in Google Colab environment with using GPU of 12.69 GB. In order to compare three methods were used Hugging Face transformers, which is provide API to quickly install and use models on selected texts.

As a test dataset were used CNN/Daily Mail Dataset, which is English language dataset containing more than 300k news articles written by journalists from CNN and the Daily Mail. However, this experiment was limited with 30 examples, due to the computational power.

For evaluation of summarization methods were used Rouge metric, which is stands for Recall-Oriented Understudy for Gisting Evaluation. Moreover, for each model were counted computation time during the summarization process. Table 1 show the results of each method tested on CNN/Daily Mail Dataset. PEGASUS provide the highest rouge scores among all methods. However, to summarize it needs 3 times more seconds than other two.

Model	Rouge-1	Rouge-2	Rouge-L	Time (seconds)
PEGASUS	0.34	0.14	0.24	3352
BART	0.24	0.12	0.22	1516
ProphetNet	0.28	0.1	0.20	1004

Table 1. Average score of rouge measures on 30 examples of CNN/Daily Mail Dataset and time spent during summarization



	highlights	pegasus_summary	bart_summary	prophetnet_summary
0	James Best, who played the sheriff on "The Dukes of Hazzard," died Monday at 88 .\n"Hazzard" ran from 1979 to 1985 and was among the most popular shows on TV .	"I laughed and learned more from Jimmie in one hour than from anyone else," co-star John Schneider says.<n>Best died of complications from pneumonia, a longtime friend and Hollywood colleague says.<n>Best was best known for his role as bumbling sheriff Rosco P. Coltrane on "The Dukes of Hazzard"	James Best was best known for his portrayal of bumbling sheriff Rosco P. Coltrane on "The Dukes of Hazzard" He died in hospice in Hickory, North Carolina, of complications from pneumonia. "I laughed and learned more from Jimmie in one hour than from anyone else in a whole year," co-star John Schneider says.	best died in hospice in hickory, north carolina, of complications from pneumonia. [X_SEP] he played bumbling sheriff rosco p. coltrane on " the dukes of hazzard "
1	A lawyer for Dr. Anthony Moschetto says the charges against him are baseless .\nMoschetto, 54, was arrested for selling drugs and weapons, prosecutors say .\nAuthorities allege Moschetto hired accomplices to burn down the practice of former associate .	Dr. Anthony Moschetto is charged in what authorities say was a failed scheme to have another physician hurt or killed.<n>Two other men -- identified as James Chmela, 43, and James Kalamaras, 41 -- were named as accomplices.<n>Police officers allegedly discovered approximately 100 weapons at Moschetto's home.	Dr. Anthony Moschetto faces criminal solicitation, conspiracy, burglary, arson, criminal prescription sale and weapons charges. He was released after posting $2 million bond and surrendering his passport. Two other men were named as accomplices, according to prosecutors. "None of anything in this case has any evidentiary value," attorney Randy Zelin says.	dr. anthony moschetto is charged in a failed scheme to have another physician hurt or killed. [X_SEP] he faces criminal solicitation, conspiracy, burglary, arson, criminal prescription sale and weapons charges.
2	"No challenge poses more of a public threat than climate change," the President says .\nHe credits the Clean Air Act with making Americans "a lot" healthier .	CNN's John Sutter talks with President Obama about climate change and public health.<n>Obama credits the Clean Air Act with making Americans "a lot" healthier.<n>Sutter: "No challenge poses more of a public threat than climate change"	President Obama says climate change is a public health issue that affects all of us. Obama: "No challenge poses more of a public threat than climate change" Obama encourages ordinary citizens, doctors and nurses to start putting some pressure on elected officials. Obama did not appear particularly concerned about the current Supreme Court challenge to the Affordable Care	president barack obama took part in a roundtable discussion this week on climate change. [X_SEP] obama sat down with cnn's john sutter for a one - on - one interview.

Figure 4. Results of each method

Results showed that all methods give approximately same results. However, bart uses more words to summarize article than other two.

III. Related works

During the research were found only two works related to text summarization for Russian and Kazakh languages. In 2020, Russian researchers published their work, where they presented the first dataset for summarization of Russian news – Gazeta. Moreover, they adapted the mBART model for text summarization of text on Russian language [4]. In 2019 researchers from Nazarbayev University published work named «Kazakh text summarization using fuzzy logic». In this work, they collected data from the different internet resources in the Kazakh language, and based on fuzzy logic made an automatic text summarization method [5].

IV. Conclusion

Finally, I found that there is no abstractive text summarization algorithms for russian and kazakh texts. One of the reason is that abstractive text summarization method is hard to realize. However, big Tech companies already created such a methods for english texts. During the

comparison of those methods, I noticed that PEGASUS showed better rouge score results, and according to paper achieved state-of-the-art results with only 1000 examples, which is conveniently for my further work. In the future, PEGASUS method is planned to be implemented for Kazakh and Russian languages in order to achieve state-of-the-art results in those languages.

References

1. Mike Lewis, Yinhan Liu, Naman Goyal, Marjan Ghazvininejad, Abdelrahman Mohamed, Omer Levy, Ves Stoyanov, Luke Zettlemoyer: BART: Denoising Sequence-to-Sequence Pre-training for Natural Language Generation, Translation, and Comprehension. arXiv preprint arXiv:1910.13461 (2019)
2. Jingqing Zhang, Yao Zhao, Mohammad Saleh, Peter J. Liu: PEGASUS: Pre-training with Extracted Gap-sentences for Abstractive Summarization. arXiv preprint arXiv:1912.08777 (2020)
3. Weizhen Qi, Yu Yan, Yeyun Gong, Dayiheng Liu, Nan Duan, Jiusheng Chen, Ruofei Zhang, Ming Zhou: ProphetNet: Predicting Future N-gram for Sequence-to-Sequence Pre-training. arXiv preprint arXiv:2001.04063 (2020)
4. Ilya Gusev: Dataset for Automatic Summarization of Russian News. arXiv preprint arXiv:2006.1106 (2020)
5. Altanbek Zulkhazhav, Zhanibek Kozhirbayev, Zhandos Yessenbayev, Altynbek Sharipbay: Kazakh Text Summarization using Fuzzy Logic.