The Problem of Understanding the Content of a Complex Question by BERT Model Construction

Vladimir Kopeyko
Belarusian State University
Minsk, Belarus
volodyakopeyko@gmail.com

Alexander Nedzved
United Institute of Informatics Problems
Minsk, Belarus
Anedzved@bsu.by

***Abstract*—This paper describes the application of the BERT natural language processing model to solve the problem of understanding the content of a complex question. In this paper there are description of the model itself, basic concepts of solution and model for training for classification. Common task are represented by example of the using of this model for practical analysis of the data. Quality results is based on preprocessing such data.**

***Keywords*—natural language processing, automated answer, language analysis, BERT model**

I. Introduction

The task of automated answering of the question is one of the important tasks for natural language analysis and processing. Modern methods for its solution allows to take good results for this task, but they are quite accurate only for limited sets of clear answers. There are problems with answers to questions that involve the expression of a subjective opinion or recommendation. The structure of the questions also has some difficulty for automated understanding. In real life, question can be represented by different forms. It can consist of only a few words or a complex structure consisting of several sentences. The understanding of such questions in natural language processing tools is possible to solve by methods of machine learning[1]. One of the basic trends in natural language processing is the use of models that are trained by simple problems. In this case, machine learning methods can be applied to other tasks through a little refinement. In addition, such models are trained on large datasets. It allow to the full extended sets, to use of an extensive context sets for solving question problems in natural language processing[2]. One of the most popular models of this type at the moment is the BERT model.

II. Creation of BERT model description

In the field of computer vision, the value of transfer training has been repeatedly demonstrated by a pre-trained neural network model. This model has good results for solving a well-known problem. For example the ImageNet with using fine-tuning can serve as the basis for a new model that solves a completely different problem. In recent years, there are many publications that describe a similar methods. It can be useful in many natural language processing problems. One such pre-trained model is the BERT model[3]. BERT (Bidirectional Encoder Representations from Transformers) is a pre-trained model for solving the problems of processing natural language with open source code, developed by researchers at Google in 2018. Most of the models for solving the problems of natural language processing considered the text during the training either from left to right, or from right to left. This unidirectional approach works well in the task of generating sentences. Usually, it can predict the next word. Then add this word to the sequence and predict the next word until the sentence is not complete. But the BERT is trained by two ways (Fig. 1). It means a deeper understanding of the language context that compared to unidirectional language models. With a change in the approach to text processing, the tasks of the model trains also change. The BERT no predict the next word in a sequence. It simultaneously solves two problems: Masked LM and Next Sentence Prediction [4]. The masked LM (MLM) consist of next basis: 1) sequences of words are used as input, 2) words are replaced by the [MASK] token in each 15 procents. The model tries to predict the initial meaning of the masked words based on the context provided by the other unmasked words in the sequence. Masking of word is meaning the model looking in both directions. After it the full context of the sentence interpreted as left and right surroundings for prediction of the masking word[5][6]. An example is shown in Table 1. The word "Wimbledon" was taken as a masked word. As a result, MLM predicts "Wimbledon" as a mask with a probability of 33

The Next Sentence Prediction (NSP) is the second task of word prediction. The main task of it is prediction of the second sentence in a pair as a continuation of the first. During BERT training, the model receives pairs of sentences as input and learns to predict whether

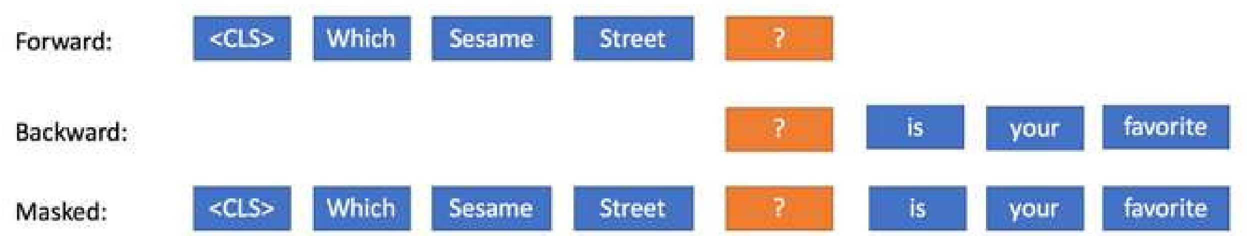

Figure 1. Example of bidirectional processing of sentence by BERT like pipeline.

Table I
EXAMPLE OF MLM ANALYSIS

Sentence	Result
Over 303,000 glasses of Pimm's are served during the [MASK] tennis championships?	32.9 Wimbledon 8.0 table 7.7 world 3.0 professional 2.3 Table

the second sentence in a pair as the next sentence in the source document. During training, 50 Thus, during

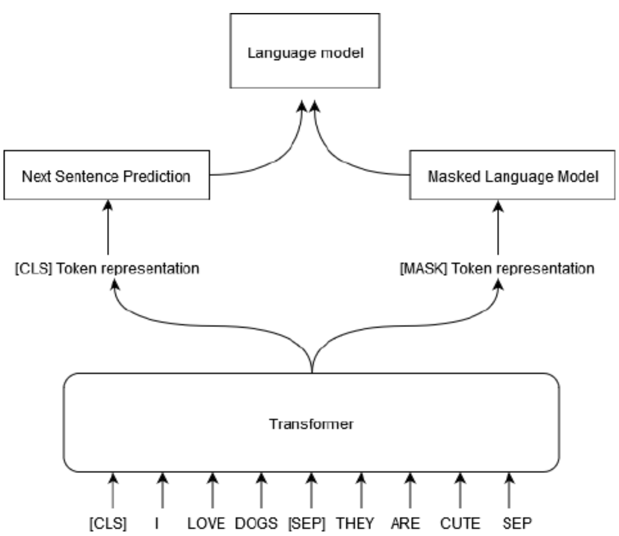

Figure 2. BERT model training scheme

training, the BERT, MLM, and NSP models are trained simultaneously to minimize the combined loss function for two tasks.

III. PREDICTION IN SENTENCE BY THE BERT MODEL

To solve the problem of improving understanding of complex question, the use of the BERT model is effective. An example of such a task can be: each question-answer pair has 30 labels of mask. Some words describe the question itself, and the other part describe the answer. Each label can have a continuous value in the range [0, 1]. At the input there is a labeled training dataset, consisting of many thousand question-answer pairs and the values of all labels for each pair. A description of pair of question and answer:

- qa id;
- question title;
- question body;
- question user name;
- question user page;
- answer
- answer user name
- answer user page
- url
- category
- host

It is necessary to determine the same labels for all question-answer pairs from the test dataset. [1] An example of a description of some tags:

- "question expect short answer" - degree of expectation of a short answer to a question;
- "question interestingness others" - degree of interest of the question for others;
- "answer helpful" - the usefulness of the answer;
- "answer level of information" - informative response;
- "answer relevance" - the relevance of the answer;
- "answer satisfaction" - satisfaction with the answer;
- "answer type instructions" - is the answer an instruction;
- "answer well written" - how well the answer is written.

There are independence of labels among themselves. Figure 3 shows the correlation matrix, where the darker the color of the cell correspond of the more closely relation of labels. It mean that those cells depend on each other. The pronounced diagonal elements that determine the dependence of the label itself. It can be seen that the greatest correlation is between the "question interestingness others and "question interestingness self" labels. Thus, the values of labels determine the degree of interest of the question for the most asking question and the degree of interest of the other question. They are most correlated in this dataset.

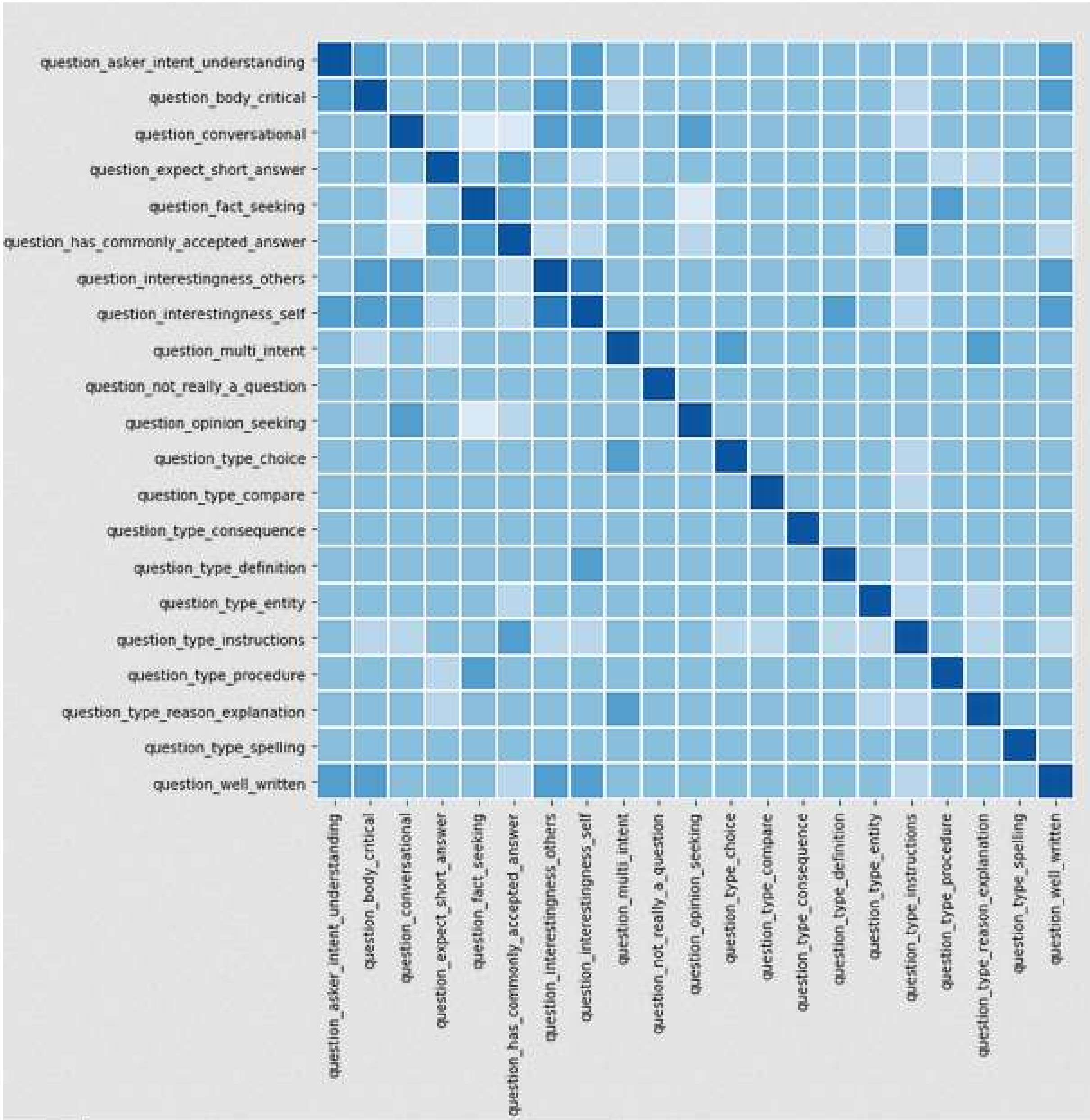

Figure 3. Label correlation matrix for test dataset.

Thus, during training, the BERT, MLM, and NSP models are trained simultaneously to minimize the combined loss function for two tasks. The solution to this problem consists of two stages: 1) training the BERT neural network, 2) its refinement (settings) by using of additional layers that needed to solve a specific task. For the first stage, a pre-trained model is used. However, it is necessary to prepare data accordingly texts of questions and answers in order of using it. The BERT is training on the NSP solution. The sentences are divided into tokens by this context. In this case, tokens are words. Then each sentence is framed by the [CLS] and [SEP] tokens, where [CLS] indicates the beginning of the sentence, [SEP] - its end. Obtained data after preliminary processing is going into the BERT model. Output data are represented as vector of size 728 elements (in the case of the usual implementation of the model) or 1024 elements (for the case of the extended version). Addition an additional layer is used as an ordinary fully connected layer. Every output vector has size 30 elements. It corresponds to a certain label, that must be predicted.

Table 2 shows examples of input pairs of questions and answers. Table 3 shows the result of using the BERT model to identify labels that explain questions

Table II
EXAMPLES OF INPUT PAIRS OF QUESTIONS AND ANSWERS

qa id	question title	question body	answer
39	Can I get back my passport after applying for a Schengen visa?	I need to travel, but the processing time takes up to 15 days according to the website.	Out of first hand experience in two different EU consulates (Italian and Greek), they actually return the passport to you and on the visa issuance date they will ask for it and they will post it there...
46	Why was Ryuk tied to Light's Death Note?	In How to Use: XIII, in the second point it says The god of Death always remains with the owner of the Death Note. indicating that a Shinigami will remain with the owner of their Death Note	According to How to use II: The owner of the note can recognize the image and voice of the original owner, i.e. a god of death. Therefore the only explanation for Light being able to see Ryuk

Table III
IDENTIFY LABELS EXPLAIN QUESTIONS AND ANSWERS IN BERT

qa id	question asker intent understanding	question body critical	question conversational	question except short answer	question fact seeking	question has commonly accepted answer	question interestingness others
39	0.927	0.563	0.192	0.53	0.687	0.556	0.658
46	0.899	0.58	0.006	0.77	0.788	0.926	0.569

and answers in the task of automatically understanding complex questions.

IV. CONCLUSIONS

The BERT model is the best solution for solving natural language processing problems by machine learning algorithms. Its characterised by availability and quick setup. It allows the use of this model for a wide range of practical applications. In addition, the pre-trained model include about 2.5 billion words. It is also an important advantage of using this model for practical purposes. The obtained results indicate applicabilition of this model for solving of problem of complex question construction.

REFERENCES

[1] Z. Chen, H. Zhang, X. Zhang, and L. Zhao Quora question pairs. 2018, Re-trieved from http://static.hongbozhang.me/doc/Quora.pdf
[2] Minh-Thang Luong, Christopher D. Manning, Achieving Open Vocabulary Neural Machine Translation with Hybrid Word-Character Models. 2016, Re-trieved from https://arxiv.org/pdf/1604.00788.pdf
[3] Jacob Devlin, Ming-Wei Chang, Kenton Lee BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. 2019, Re-trieved from https://arxiv.org/pdf/1810.04805.pdf
[4] Kevin Clark, Urvashi Khandelwal, Omer Levy Manning What Does BERT Look At? An Analysis of BERT's Attention. 2019, Re-trieved from https://arxiv.org/pdf/1906.04341.pdf
[5] Xing Wu, Tao Zhang, Liangjun Zang, Jizhong Han, Songlin Hu, Mask and Infill : Applying Masked Language Model to Sentiment Transfer. 2019, Re-trieved from https://www.ijcai.org/proceedings/2019/0732.pdf
[6] Noah A. Smith Contextual Word Representations: A Contextual Introduction. 2019, Re-trieved from https://arxiv.org/pdf/1902.06006.pdf
[7] Yifu Sun, Haoming Jiang, Contextual Text Denoising with Masked Language Models. 2019, Re-trieved from https://arxiv.org/pdf/1910.14080.pdf
[8] Wei Shi, Vera Demberg Next Sentence Prediction helps Implicit Discourse Relation Classification within and across Domains. 2019, Re-trieved from https://www.aclweb.org/anthology/D19-1586.pdf
[9] Felipe Alfaro, Marta R. Costa-jussà, José A. R. Fonollosa BERT Masked Language Modeling for Co-reference Resolution. 2019, Re-trieved from https://www.aclweb.org/anthology/W19-3811.pdf
[10] Jan Christian Blaise Cruz Evaluating Language Model Finetuning Techniques for Low-resource Languages. 2019, Re-trieved from https://arxiv.org/pdf/1907.00409.pdf
[11] Sneha Chaudhari, Gungor Polatkan An Attentive Survey of Attention Models. 2019, Re-trieved from https://arxiv.org/pdf/1904.02874.pdf
[12] Ashish Vaswan, Noam Shazeer, Niki Parmar, Jakob Uszkoreit Attention Is All You Need. 2017, Re-trieved from https://arxiv.org/pdf/1706.03762.pdf

Использование BERT модели в задаче понимания содержания сложного вопроса

Копейко В.С., Недзьведь А.М.

Данная работа посвящена описанию применения модели обработки естественного языка BERT для решения задачи понимания содержания сложного вопроса. Сначала вводится описание самой модели, ее основные концепции и задачи, на которых модель изначально тренируется. Затем рассматривается пример использования этой модели в практических целях, анализ исходных данных и описание процесс их предобработки.

Received 14.01.2020