УДК 004

А. В. Бурлаченко
Л. В. Ібрагімова

DEEP LEARNING MODEL (NLP) – BERT

Вінницький національний технічний університет

Анотація *У статті розповідається про NLP моделі та описується така модель машинного навчання як BERT.*
Ключові слова: машинне навчання, глибоке навчання, обробка природної мови (NLP).

Abstract *The article defines the concept of natural language processing and describes the BERT model.*
Keywords: machine learning, deep learning, natural language processing (NLP).

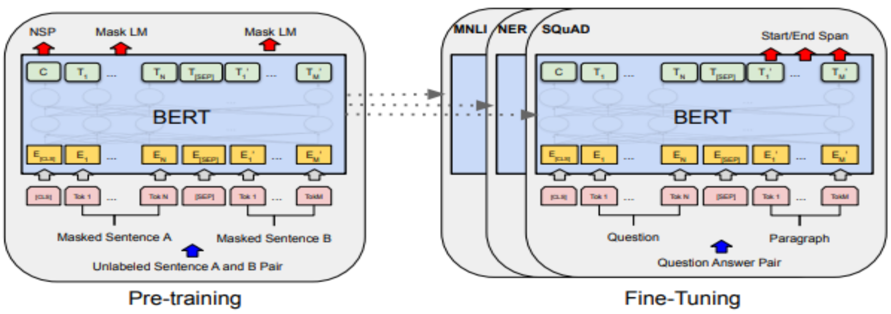
Natural language processing (NLP) refers to the branch of computer science — and more specifically, the branch of artificial intelligence or AI — giving computers the ability to understand text and spoken words in much the same way as human beings can.

NLP combines computational linguistics, rule-based modeling of human language with statistical, machine learning, and deep learning models. Together, these technologies enable computers to process human language in the form of text or voice data and to “understand” its full meaning along with the speaker or writer’s intent and sentiment.

Language model pre-training has been shown to be effective for improving many natural language processing tasks (Dai and Le, 2015; Peters et al., 2018a; Radford et al., 2018; Howard and Ruder, 2018). These include sentence-level tasks such as natural language inference (Bowman et al., 2015; Williams et al., 2018) and paraphrasing (Dolan and Brockett, 2005), the aim of which to predict the relationships between sentences by analyzing them holistically, as well as token level tasks such as to name entity recognition and answer the questions where models are required to produce fine-grained output at the token level.

BERT stands for Bidirectional Encoder Representations from Transformers. Unlike recent language representation models (Peters et al., 2018a; Radford et al., 2018), BERT is designed to pretrain deep bidirectional representations from unlabeled text by conditioning jointly on both left and right context in all layers. As a result, the pre-trained BERT model can be fine-tuned with just one additional output layer to create state-of-the-art models for a wide range of tasks such as question answering and language inferencing without substantial task specific architecture modifications.

BERT is conceptually simple and empirically powerful. It obtains new state-of-the-art results on eleven natural language processing tasks including pushing the GLUE score to 80.5% (7.7% point absolute improvement), MultiNLI accuracy to 86.7% (4.6% absolute improvement), SQuAD v1.1 question answering Test F1 to 93.2 (1.5 point absolute improvement) and SQuAD v2.0 Test F1 to 83.1 (5.1 point absolute improvement).



Input/Output: To make BERT handle a variety of down-stream tasks our input representation is able to unambiguously represent both a single sentence and a pair of sentences (e.g., h Question, Answeri) in one token sequence. Throughout this work, a “sentence” can be an arbitrary span of contiguous text rather than an actual linguistic sentence. A "sequence" refers to the input token sequence to BERT, which may be a single sentence or two sentences packed together. We use WordPiece embeddings (Wu et al., 2016) with a 30,000 token vocabulary. The first token of every sequence is always a special classification token ([CLS]). The final hidden state corresponding to this token is used as the aggregate sequence representation for classification tasks. Pairs of sentence are packed together into a single sequence. We differentiate the sentences in two ways. First, we separate them with a special token ([SEP]). Second, we add a learned embedding to every token indicating whether it belongs to sentence A or sentence B. As shown in Figure 1, we denote input embedding as E, the final hidden vector of the special [CLS] token as C ∈ R H, and the final hidden vector for the i th input token as Ti ∈ R H. For a given token, its input representation is constructed by summing the corresponding token, segment, and position embeddings. A visualization of this construction can be seen in Figure 2.

Conclusion: Recent empirical improvements due to transfer learning with language models have demonstrated that rich, unsupervised pre-training is an integral part of multiple language understanding systems. In particular, these results enable even low-resource tasks to benefit from deep unidirectional architectures. Our major contribution is further generalizing these findings to deep bidirectional architectures, allowing the same pre-trained model to successfully tackle a broad set of NLP tasks.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ:

1. Визначення і основні поняття – Режим доступу до ресурсу: https://www.ibm.com/cloud/learn/natural-language-processing
2. BERT model – Режим доступу до ресурсу: https://arxiv.org/pdf/1810.04805.pdf -

Бурлаченко Артем Віталійович – ст. гр. СА-20б факультет інтелектуальних інформаційних технологій та автоматизації, Вінницький національний технічний університет, м. Вінниця, e-mail: simpleprofi@gmail.com

Науковий керівник: ***Ібрагімова Людмила Володимирівна*** — старший викладач кафедри іноземних мов, Вінницький національний технічний університет, м. Вінниця, e-mail: milatvin@ukr.net

Burlachenko Artem - Faculty of Intelligent Information Technologies and Automation, Vinnytsia National Technical University, Vinnytsia, e-mail: simpleprofi@gmail.com

Ibrahimova Liudmyla V. — Senior Lecture, Chair of Foreign Languages, Vinnytsia National Technical University, Vinnytsia, e-mail: milatvin@ukr.net