2019 IEEE International Conference on Consumer Electronics
- Taiwan (ICCE-TW)

Learning Text Representations for Finding Similar Exercises

Mengfei Feng, Yishuai Chen, Yuchun Guo, Yongxiang Zhao
School of Electronics and Information Engineering
Beijing Jiaotong University, Beijing, China 100044
Email: {17120052, yschen, ychguo, yxzhao}@bjtu.edu.cn

Guowei Fu
TAL Education Group Inc.
Beijing, China, 100080
Email: fuguowei@100tal.com

***Abstract*—Mathematical Intelligent Tutor System brings great convenience for both teachers and students. A basic task in the system is to find similar exercises, which examine students the same skills or knowledge. Inspired by previous work, we propose a new model called Siamese based Bidirectional Encoder Representations from Transformer (SBERT). After training on our Chinese math exercises dataset, AUC(Area Under Curve) of SBERT model can reach up to 0.90, which is higher than that of existed models. Visualization analysis also proves that our model obtains better text representing performance of exercises than previous work.**

***Index Terms*—Similar exercises, Text representations**

I. Introduction

Nowadays Mathematical Intelligent Tutor System (MITS) brings great convenience for both teachers and students. When a student answer an exercise wrongly in MITS, it is common to recommend similar exercises, which examine the student the same skills or knowledge. So finding similar exercises (FSE) is a basic and essential task in MITS.

Great progress in Natural Language Processing (NLP) field provides promising approaches for FSE. However, because of the complex mathematical logic and relationships, there still exists a gap between an exercise's text description and process of identifying knowledge point. In detail, Bidirectional Encoder Representations from Transformer (BERT) [1] involves the interactions between words, and may help us capture the mathematical logic and relationships in texts of math exercises. Other than that, Siamese architecture [2] can be used for learning appropriate exercises' vector-space representations, and Siamese based Long Short-Term Memory (SLSTM) model [3] performs well on FSE task. In addition to FSE task, SLSTM model is also used in Paraphrase Identification task [4].

Although BERT is a general model on many NLP tasks, it has not been proved that BERT can focus on mathematical characteristics on FSE task. Furthermore, it simply separates dissimilar exercises, rather than focusing on increasing distances between dissimilar exercises while reducing distances between similar exercises. Siamese architecture can learn exercises' vector-space representations, in which the positions of similar exercises are close and those of dissimilar exercises are far. Previous Siamese work used LSTM to learn exercises' representations on FSE task [3]. However, LSTM is less effective than Transformer used in BERT [1]. Based on that, we apply BERT in Siamese architecture. In this way, both BERT and Siamese architecture can overcome each other's shortcomings and leverage their own advantages. Our contributions are two folds:

- We propose a new model called SBERT based on Siamese network and BERT. It is observed that, after training on our Chinese math exercises dataset, AUC of SBERT can reach up to 0.90, which is 0.32 and 0.05 absolutely higher than that of BERT and SLSTM models, respectively;
- We conduct visualization analysis of exercises' representations obtained by SBERT, proving SBERT obtains better text representing performance than existing models.

II. SBERT Model

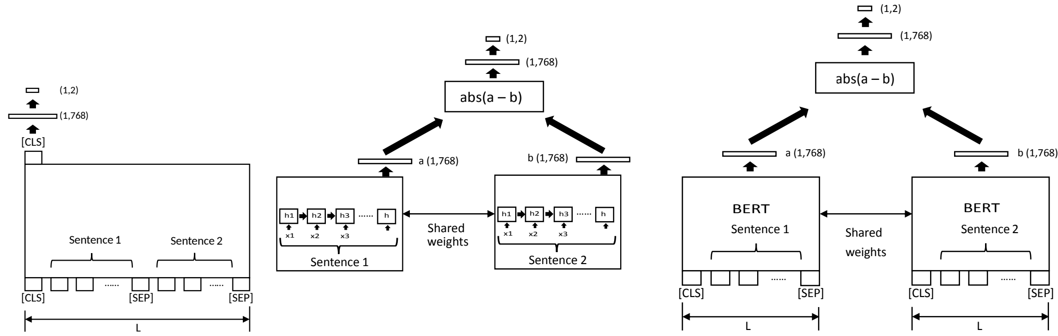
A brief introduction of the proposed model is to be shown in this section. We first introduce BERT [1] and SLSTM [3], [4]. As is shown in Fig. 1(a), when BERT is used for identifying similarity, it simply concatenates two sentences' texts as input. The end of a sentence is denoted by *[SEP]*. After representing each position of the input as a vector by using bidirectional Transformer, BERT leverages the hidden vector corresponding to the starting position *[CLS]* to predict whether the two sentences are similar. Fig. 1(b) shows how SLSTM identifies similarity. After sending two sentences to two LSTM models with the same structure and weights, the absolute value of the result, which is obtained by substracting the two exercises' text representations, is sent to a full-connection network to determine whether they are similar.

Based on them, we propose the Siamese based Bidirectional Encoder Representations from Transformer (SBERT) model, which is shown in Fig. 1(c). Our SBERT model captures mathematical logic and relationships by using BERT and learns appropriate vector-space representation by using Siamese Network. The following two methods are adopted:

- SBERT-C: Same with BERT, we use the vector corresponding to the starting position *[CLS]* as the representation of an exercise.
- SBERT-M: Different from BERT, we average vectors corresponding to all the positions as the representation of an exercise.

III. Performance Evaluation

This section presents the performance evaluation results.



(a) BERT (b) SLSTM (c) SBERT

Fig. 1: Model Structure

A. Data

We evaluate our model based on a real Chinese math exercises dataset, which is currently used in TAL Education Group Inc., one of the largest online education companies in China. In the dataset, each exercise has a tag, e.g., Dot Product, Geometric Series, etc. Similar to [3], we generate exercise text pairs with labels marking whether they are similar, i.e., whether they are under the same tag. Finally, we obtain 8125 exercise text pairs with a 6500/812/813 training/validating/testing split.

B. Settings

We use BERT and SLSTM models as baselines. We denote a threshold of length as L and truncate exercises at the L^{th} tokens. We set L to 128 for SLSTM and SBERT models as the mean length of all exercises is about 128. As BERT concatenates two sentences to accomplish the classification task, we set L to 256. The dimension of hidden units is set to 768, which is the same with [1].

C. Results

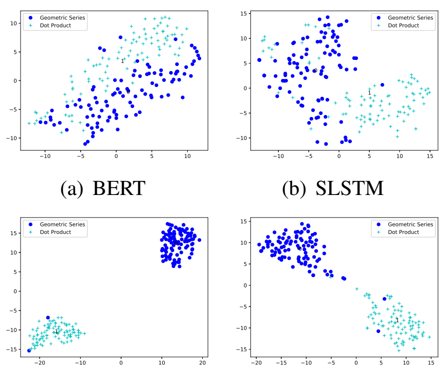
The results of similarity identification are shown in Table I. The AUC of SBERT-C model is 0.88, which is 0.03 absolutely higher than that of SLSTM model, meaning that BERT is indeed more powerful than LSTM in obtaining the representations of math exercises. It is also 0.30 absolutely higher than that of BERT, meaning that Siamese architecture is more suitable for identifying similarity. Furthermore, the AUC of SBERT-M model is 0.90, which is 0.32 and 0.05 absolutely higher than that of BERT and SLSTM models, respectively. It is also 0.02 absolutely higher than that of SBERT-C model, suggesting that the mean operation can improve the model's performance since it involves more information.

D. Visualization Analysis

We finally conduct visualization analysis of the obtained exercises' representations. For FSE task, it is important to learn text representations in which similar exercises are closer while dissimilar exercises are farther. To show the results intuitively, we first select exercises under two tags that are randomly selected (Geometric Series and Dot Product are selected in Fig 2), and then reduce the dimension of obtained representations by t-SNE. The positions of the processed representations are finally shown in Fig 2. Compared with BERT and SLSTM models, similar exercises are closer and dissimilar exercises are farther by using SBERT models. Thus, the SBERT models have better text representing performance.

TABLE I: Performance of Different Models

Model	AUC	Accuracy	F1(macro)
BERT	0.58	0.56	0.56
SLSTM	0.85	0.77	0.77
SBERT-C	0.88	0.82	0.83
SBERT-M	**0.90**	**0.84**	**0.83**



(a) BERT (b) SLSTM

(c) SBERT-C (d) SBERT-M

Fig. 2: Representation Visualization

Acknowledgement

This work was supported by National Natural Science Foundation of China under Grants 61572071, 61872031, and 61301082.

References

[1] Devlin J, Chang M, Lee K, et al. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding.[J]. arXiv: Computation and Language, 2018.

[2] Chopra S, Hadsell R, Lecun Y, et al. Learning a similarity metric discriminatively, with application to face verification[C]. computer vision and pattern recognition, 2005: 539-546.

[3] Liu Q, Huang Z, Huang Z, et al. Finding Similar Exercises in Online Education Systems[C]. knowledge discovery and data mining, 2018: 1821-1830.

[4] Mueller J, Thyagarajan A. Siamese recurrent architectures for learning sentence similarity[C]. national conference on artificial intelligence, 2016: 2786-2792.