Jointly Considering Siamese Network and MatchPyramid Network for Text Semantic Matching

To cite this article: Zhoulin Tang and Jian Li 2019 *IOP Conf. Ser.: Mater. Sci. Eng.* **490** 042043

IOP Conf. Series: Materials Science and Engineering **490** (2019) 042043 doi:10.1088/1757-899X/490/4/042043

Jointly Considering Siamese Network and MatchPyramid Network for Text Semantic Matching

Zhoulin Tang[1,*], Jian Li[2]

[1]School of Computer Science, Beijing University of Posts and Telecommunications, Beijing, China

*Corresponding author e-mail: zhoulin.tang@bupt.edu.cn, lijian@bupt.edu.cn

Abstract. When users ask questions in the Q&A website, the Q&A website often recommends similar questions to users. In order to improve the accuracy of recommendation, we propose a new text matching model: a text matching model based on Siamese semantic network and MatchPyramid model. Specifically, the algorithm combines the advanced features of Siamese semantic network with MatchPyramid model, and use these advanced features for classification. We compare our proposed model with several other models on the dataset provided by Quora and Alibaba. The results show that our proposed model performs better than other models.

1. Introduction

Text semantic matching is an important task. Recognizing semantic similarity between sentences has different applications, such as machine translation[1], question and answering[2], natural language understanding and document retrieval[3].

Given two short texts $T_1 = (w_1, w_2, ..., w_m)$ and $T_2 = (v_1, v_2, ..., v_n)$, the formalized representation of the two text matches is as follows:

$$match(T_1, T_2) = F_1(\Phi_1(T_1), \Phi_1(T_2)) \tag{1}$$

or:

$$match(T_1, T_2) = F_2(\Phi_2(T_1, T_2)) \tag{2}$$

w_i and v_j represent the words in T_1 and T_2, respectively. Φ_1 represents a model that converts a single sentence into a feature vector, while Φ_2 represents a model that converts two sentences into a feature vector. F_1 is used to measure similarity between two feature vectors, while F_2 a is used to represent a two-category model or a regression model for predicting similarity of sentences.

Problems such as synonym, word spelling and sentence order make it difficult to judge whether sentences are similar or not. Word-based comparison is a method of simple similarity detection. Firstly, we can use TF-IDF or bag of words to represent sentences as vectors. Then, the cosine distance between vectors or other metrics can be calculated to express the similarity of sentences. However, these methods ignore the order of the words in the sentence, and can not overcome the sparsity problem. Besides, these methods also raise a problem that the dimension of feature vector is too high.

Deep text matching models[4][20] show better results on semantic similarity, it can not only learn various matching patterns between sentences but also overcome the problems which may occur in the TF-IDF and bag of words algorithms. Models of deep text matching include Siamese Recurrent Neural Network(RNN)[5], Siamese Convolutional Neural Network(CNN)[6] and MatchPyramid network[7]. However, single MatchPyramid model concentrate more on the connection between two sentences, while single Siamese semantic model consider more about the context information of the single sentence. Therefore, this paper proposes a text semantic matching model combining Siamese Network and MatchPyramid network.

2. Model introduction

2.1. Text matching based on Siamese Network

Siamese network contains two identical sub-networks. The word 'Identical' means the each sub-network has the same structure and parameters. The sub-networks can be CNN[8], RNN[9], Long Short Term Memory networks(LSTM)[10], Gated recurrent units(GRU)[11] or other networks. This structure has two inputs which corresponding to the two sentence vectors used to compare the similarities and one result. The identical sub-networks can extract the features of two sentences which can be used to compare the semantics similarity of the two input sentences.

The Siamese neural network based on RNN is shown in Fig. 1, x_1 and x_2 are the inputs of Siamese RNN. h_1 and h_2 represent RNN. '*classify*' represents the model for similarity evaluation by using the features extracted from the Siamese neural network. This model can be traditional machine learning algorithms such as Support Vector Machines(SVM), Random Forests(RF), Logistic Regression(LR) or a fully connected network.

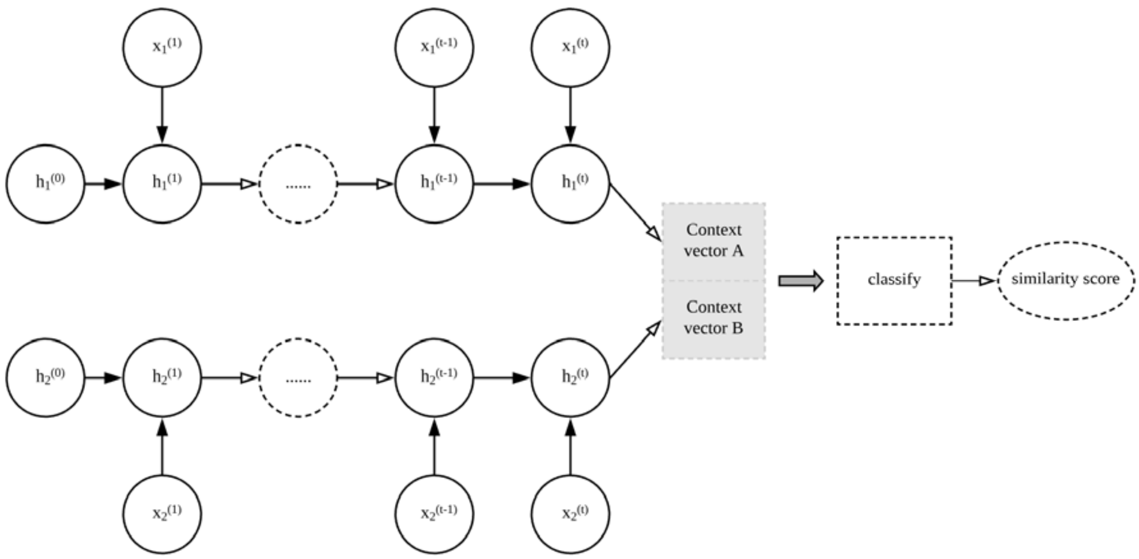

Figure 1. Siamese neural network based on RNN

2.2. Text matching based on MatchPyramid model

Deep text matching model needs to get the rich match patterns in text semantic matching process. Taking text semantic matching as an example, here are two sentences with the same semantic meaning:

T_1: *The president of the United States signed the economic and trade related act.*

T_2: *Trump signed the economic and trade related bill.*

Match patterns include different levels: words, phrases and n-terms. Firstly, there are many word-level matching patterns, including identical word such as "*economic*" in both T_1 and T_2, and similar word such as "*act*" in T_1 and "*bill*" in T_2. And semantic n-term matching between "*president of the United States*" in T_1 and "*Trump*" in T_2.

The MatchPyramid model is shown in Fig. 2. MatchPyramid neural network construct word-level similarity matrix to catch structures mentioned above. The dot product between the word vectors of word w_i and word v_j constitutes the similarity matrix. So similarity matrix consider the relation between two sentences on word-level. It is a strong supplement to the context vector feature generated by Siamese neural network which only consider the single sentences information but ignore the link between two sentences.

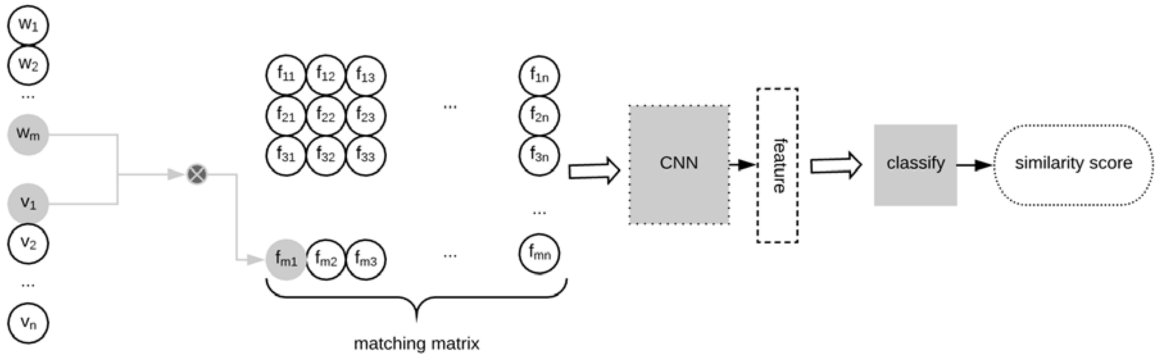

Figure 2. MatchPyramid

3. Joint Model

3.1. Model framework

The MatchPyramid model consider more about the context information of the single sentence, while the Siamese semantic model concentrate more on the connection between two sentences, inspire by this idea, we propose a new model that jointly considering Siamese network[19] and MatchPyramid network for text semantic matching in two different aspects, the single sentence context feature and the connection feature between sentences. Model structure is shown in Fig. 3.

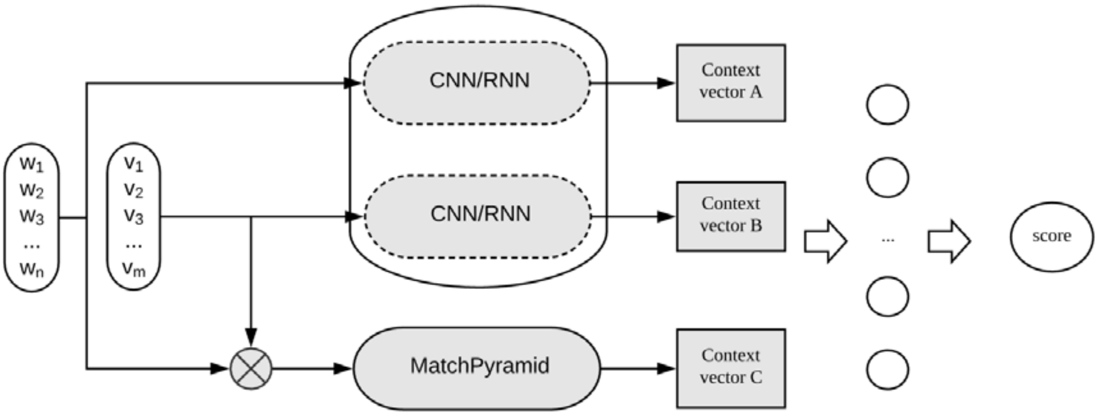

Figure 3. Joint model

3.2. Embedding layer

The embedding layer uses a method of word embedding, which is a type of method that uses dense vector representations to represent words and documents. Word embedding is an improvement of the traditional coding scheme. The traditional method uses a large and sparse vector to represent each word. The dimension of each vector is equal to the size of the vocabulary, such as bag of words and TF-IDF algorithms.

In the embedding layer, each input word will be transferred into a feature vector. Each feature vector can be initialized randomly or by using pre-trained Glo2Ve[12] and word2vec[13] vectors，In our experiments we combined two initialization methods to initialize each vector.

3.3. Feature extraction

As demonstrated in Fig 3, we combine the high-order features extracted from the MatchPyramid model with the context vector generated by Siamese semantic network, then use these combined features to classify.

In the experiments, two kinds of sub-networks of Siamese neural networks are used: CNN and LSTM. Weights and parameters are shared between sub-networks.

IOP Conf. Series: Materials Science and Engineering **490** (2019) 042043 doi:10.1088/1757-899X/490/4/042043

The elements in the similarity matrix of the MatchPyramid model use the inner product between words embedding, and then use CNN for feature extraction.

3.4. Classification method and loss function

When we extract all the features, we use full-connected network to classify the samples and use the cross entropy function as the loss function of the network. The formulation of cross entropy function as follow:

$$-\sum_{i=1}^{n} y_i \log f(x_i,\theta) + (1-y_i)\log(1-f(x_i,\theta)) \tag{3}$$

y_i is the sample label, x_i stands the sentence, and θ is the parameter of our model.

4. Data and experiment

4.1. Experiment data

The experimental data comes from Quora[1] and Ant Financial[2]. The data distributions of the Quora and Ant Financial data sets are shown in Tables 1.

Table 1. Data overview

Data sets	Positive sample	Negtive sample	Total sample	Sample proportion
Quora	149263	255027	404290	0.58
Ant Financial	18685	83792	102477	0.22

After preprocessing, we randomly extracted 10000 samples from Quora dataset and 5000 samples from Ant Financial dataset as test set. The rest part of two dataset are used as training set.

4.2. Data preprocessing

Stanford University's symbolic processing tool is used in the preprocessing procedure. We firstly remove the stop words and then delete the samples whose length is less than 10 characters. Wikipedia's English corpus is used to train Word2vec and Glo2Ve word vectors, and the dimension of word vectors is fixed at 100 in all experiments.

The dataset provided by Ant Financial is an Chinese dataset. we use Jieba(Chinese text segmentation) to segment words. In order to get better results, some scientific and technological vocabulary is added to the vocabulary.

4.3. Experiments and Results

In order to evaluate our model, we set several baseline by using different models. We believe that different models lead to different results. So we tried several different structures based on the Siamese neural network and sub-model for extract features, including bag of words, doc2vec[14], CNN and LSTM. Besides, we also use the MatchPyramid model to be another baseline.

The experimental environment is Python3.5. The model is built under the TensorFlow framework. After the preprocessing process, we utilize single GPU1080TI to train the model. The accuracy and recall score are used as the evaluation metric of the experiment. Table 2, Table 3 shows the results of our experiments on Quora dataset and Ant Financial dataset.

1 http://qim.ec.quoracdn.net/quora_duplicate_questions.tsv

2 https://dc.cloud.alipay.com/index#/topic/data?id=3

Table 2. performance of models on Quora data set

Methods	accuracy(%)	recall(%)
Siamese with bag of words	76.3	80.2
Siamese with doc2vec	77.1	77.0
Siamese with LSTM	83.2	83.8
Siamese with CNN	83.5	83.1
MatchPyramid	83.1	82.9
Siamese with CNN + MatchPyramid	**83.6**	**83.0**
Siamese with LSTM + MatchPyramid	**85.9**	**84.4**

Table 3. performance of models on Ant Financial data set

Methods	accuracy(%)	recall(%)
Siamese with bag of words	78.3	84.2
Siamese with doc2vec	79.6	81.7
Siamese with LSTM	82.2	82.8
Siamese with CNN	83.5	83.1
MatchPyramid	83.0	83.0
Siamese with CNN + MatchPyramid	**83.2**	**85.0**
Siamese with LSTM + MatchPyramid	**87.9**	**87.4**

4.4. Analysis

From the above results, it can be seen that the Siamese neural network based on LSTM and the MatchPyramid model are the best. LSTM extracts the context information of single sentence, and the MatchPyramid model extracts the interaction information between two sentences which makes the extracted information more abundant and results in higher accuracy score. According to the results in the table, the validity of our model can be proved.

5. Conclusion and Future Work

In the new text semantic matching model mentioned above, the model combines the MatchPyramid model and the Siamese neural network model, and finally generates a matching score which can predict whether the sentences pair has the same meaning or not. The model considers features in two different aspects, single sentences context feature and the connection feature between sentence pair. In semantic matching, the semantic similarity of two sentences can be considered as one sentence is translated by another sentence, so in the future work, the translation model[16][17] can be used to do text semantic matching, for example, using the attention model[18].

References

[1] Bahdanau D, Cho K, Bengio Y. Neural Machine Translation by Jointly Learning to Align and Translate[J]. Computer Science, 2014.
[2] Unger C, Freitas A, Cimiano P. An Introduction to Question Answering over Linked Data[C]// Reasoning Web International Summer School. Springer, Cham, 2014:100-140.
[3] Kang I H, Kim G C. Query type classification for web document retrieval[C]// Proc. International ACM SIGIR Conference on Research and Development in Informaion Retrieval. 2003:64-71.
[4] Guo J, Fan Y, Ai Q, et al. A Deep Relevance Matching Model for Ad-hoc Retrieval[C]// ACM International on Conference on Information and Knowledge Management. ACM, 2016:55-64.
[5] Neculoiu P, Versteegh M, Rotaru M. Learning Text Similarity with Siamese Recurrent Networks[C]// Repl4nlp Workshop at ACL. 2016.
[6] Low K B, Sheikh U U. Learning hierarchical representation using Siamese Convolution Neural Network for human re-identification[C]// Tenth International Conference on Digital Information Management. IEEE, 2016:217-222.

[7] Pang L, Lan Y, Guo J, et al. Text Matching as Image Recognition[J]. 2016.
[8] Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks[C]// International Conference on Neural Information Processing Systems. Curran Associates Inc. 2012:1097-1105.
[9] Li S, Wu C, Li H, et al. FPGA Acceleration of Recurrent Neural Network Based Language Model[C]// IEEE, International Symposium on Field-Programmable Custom Computing Machines. IEEE, 2015:111-118.
[10] Sundermeyer M, Ney H. From feedforward to recurrent LSTM neural networks for language modeling[M]. IEEE Press, 2015.
[11] Dang L M, Sadeghi-Niaraki A, Huynh H D, et al. Deep Learning Approach for Short-Term Stock Trends Prediction based on Two-stream Gated Recurrent Unit Network[J]. IEEE Access, PP(99):1-1.
[12] Mikolov T, Sutskever I, Chen K, et al. Distributed representations of words and phrases and their compositionality[C]// International Conference on Neural Information Processing Systems. Curran Associates Inc. 2013:3111-3119.
[13] Pennington J, Socher R, Manning C. Glove: Global Vectors for Word Representation[C]// Conference on Empirical Methods in Natural Language Processing. 2014:1532-1543.
[14] Le Q V, Mikolov T. Distributed Representations of Sentences and Documents[J]. 2014, 4:II-1188.
[15] Koushik J. Understanding Convolutional Neural Networks[J]. 2016.
[16] Yamada K, Knight K. A syntax-based statistical translation model[C]// Meeting on Association for Computational Linguistics. Association for Computational Linguistics, 2001:523-530.
[17] Bahdanau D, Cho K, Bengio Y. Neural Machine Translation by Jointly Learning to Align and Translate[J]. Computer Science, 2014.
[18] Parikh, Ankur P, Täckström, Oscar, Das, Dipanjan, et al. A Decomposable Attention Model for Natural Language Inference[J]. 2016:2249-2255.
[19] Shen Y, Wang H, Dai Y. Deep siamese network-based classifier and its application[J]. Computer Engineering & Applications, 2018.
[20] Wang L, Li Y, Huang J, et al. Learning Two-Branch Neural Networks for Image-Text Matching Tasks[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017, PP(99):1-1.