An Enhanced Convolutional Neural Network Model for Answer Selection

Jiahui Guo #1, Bin Yue #2, Guandong Xu ‡†3, Zhenglu Yang* #†4, Jin-Mao Wei #5
#College of Computer and Control Engineering, College of Software, Nankai University, Tianjin, China
†Institute of Statistics, Nankai University, Tianjin, China
‡ Advanced Analytics Institute, University of Technology Sydney, Sydney, Australia
1 2{guojiahui, yuebin}@mail.nankai.edu.cn, 3Guandong.xu@uts.edu.au
4 5{yangzl, weijm}@nankai.edu.cn

ABSTRACT

Answer selection is an important task in question answering (QA) from the Web. To address the intrinsic difficulty in encoding sentences with semantic meanings, we introduce a general framework, i.e., Lexical Semantic Feature based Skip Convolution Neural Network (LSF-SCNN), with several optimization strategies. The intuitive idea is that the granular representations with more semantic features of sentences are deliberately designed and estimated to capture the similarity between question-answer pairwise sentences. The experimental results demonstrate the effectiveness of the proposed strategies and our model outperforms the state-of-the-art ones by up to 3.5% on the metrics of MAP and MRR.

Keywords

question answering; answer selection; CNN

1. INTRODUCTION

As the explosive growth of data on the Web, it becomes more difficult to provide accurate information to users. Question answering (QA) systems, as an alternative to keyword-based search engines, can understand the natural language questions and offer exact answers concisely. Generally, QA systems have a pipeline architecture which is implemented in two major steps, i.e., candidate retrieval and answer selection. This paper focuses on answer selection, i.e., our goal is to select those correct sentences that contain the information to answer the question from a set of candidates obtained via search engines or information extraction systems.

Previous studies have mostly focused on the transformation between syntactic structures of question-answer pairwise sentences. Recently, the deep-learning based answer selection techniques [1, 4, 6, 7] validate their effectiveness to utilize semantic information fully, which generally involve two steps: (1) modeling sentence representations of the input question and the answer, based on a neural network architecture (e.g., CNN); and (2) training a binary classifier based on appropriate similarity measurements.

We propose a general framework, i.e. LSF-SCNN, which stems from the traditional CNN and employs several optimization strategies, such as lexical semantic feature (LSF), skip convolution (SC), and k-max average pooling (KMA).

*Corresponding author

WWW'17 Companion, April 3–7, 2017, Perth, Australia.
ACM 978-1-4503-4914-7/17/04.
http://dx.doi.org/10.1145/3041021.3054216

2. LSF-SCNN MODEL

The LSF-SCNN model is comprised of three modules (as depicted from the bottom to the top in Fig. 1): (1) for each question q and answer a, we build the lexical semantic features of each word to encode the correlation between q and a, which will be then combined with the word embeddings to construct a semantically richer sentence representation; (2) the sentence representations of q and a are fed into the skip convolution layer and k-max average pooling to produce the final representations X_q and X_a for the question and answer, respectively; and (3) in the classification process, X_q and X_a are used to compute the similarity score x_{sim} based on the learned similarity matrix U. After that, the combination of X_q, x_{sim} and X_a is used to train a binary classifier and predicts whether a is the correct answer for q.

2.1 Lexical Semantic Features

The existing CNN based approaches only generate the respective representations of q and a, while ignoring the correlation between them. Several studies remedy the issue to some extent, i.e., word co-occurrence count features [7], surface-form string matching [2]. We argue that these techniques are limited and LSF is proposed to generalize the semantic similarities between words in q and a by mapping them into a more fine-grained similar degree ranging in [0,t], which can be calculated as following:

$$LSF(q_i) = \lceil (1 - \max(0, \max_{1 \le j \le l_2} cos\langle \vec{w}_{q_i}, \vec{w}_{a_j} \rangle)) \times t) \rceil \quad (1)$$

For example, as shown in Fig. 1, LSF(*general*) is 3 because *chief* is the most similar word to *general* in a, and their cosine similarity 0.792 is mapped to the degree 3 (t = 10).

2.2 Skip Convolution

We introduce the skip convolution (SC) as an effective mechanism for the convolution operation. It allows the filters to convolve not only the adjacent words (i.e., continuous grams) in sequences but also the skipped words (i.e., skip-grams), and thus provides more effective features for pooling layer. Take the sentence *the cat sat on the mat* for example, the skip-grams, i.e., "cat sat on mat" and "cat on the

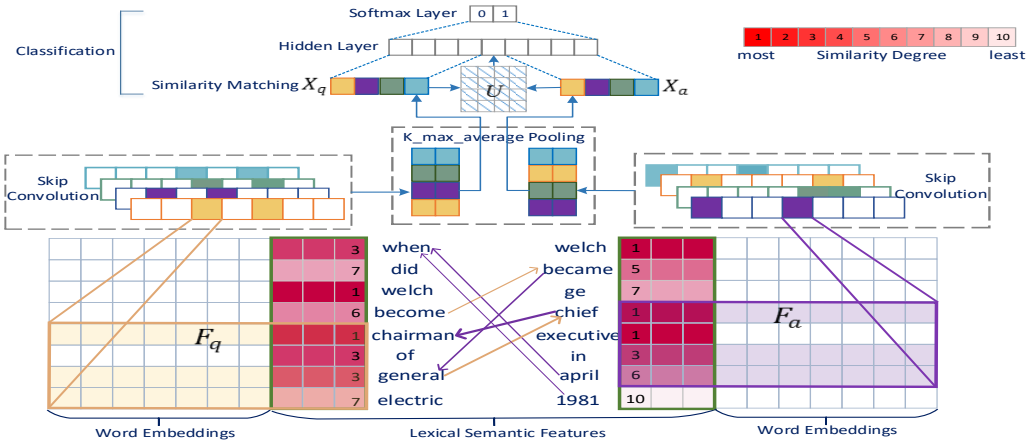

Figure 1: Lexical semantic feature based skip CNN (LSF-SCNN) model

Table 1: Our models v.s. the state-of-the-art ones

	QASent		WikiQA	
Models	MAP	MRR	MAP	MRR
Yu et al. (2014) [7]	.7113	.7846	—	—
Yin et al.(2015) [6]	—	—	.6921	.7108
Severyn et al.(2015) [1]	.7459	.8078	.6661	.6851
Wang et al.(2016) [4]	.7714	.8447	.7058	.7226
This Work (LSF-SCNN)	**.8069**	**.8796**	**.7171**	**.7336**

mat", can effectively refine the main meaning of the sentence, which indicates that they provide more effectiveness than the continuous grams employed in the previous work.

2.3 k-max Average Pooling

The pooling operation, which maps each feature map to a single value, is used to aggregate features and generate a fixed-length representation. Given its ability to remove noisy information, our proposed method k-max average pooling (KMA) outperforms the traditional methods, i.e., *average* and *max*, in terms of extracting the k highest features from the input sentences and using their average as the final pooling results for a single feature map.

3. EXPERIMENTAL EVALUATION

We conduct experiments on two datasets, i.e., *QASent* [3] and *WikiQA* [5]. We follow the same experimental setup and word embeddings as that of previous work [1, 4].

Table 1 illustrates the results of our model (i.e., parameters are tuned by using the five-fold validation on the DEV sets). We can see LSF-SCNN performs the best, and is better than the previous best one [4] by up to 3.5% on *QASent*, and 1.2% on *WikiQA* in terms of MAP and MRR. The reason for the superior of our proposed model is due to the highly effective word-level and phrase-level granularity features with the deliberately designed mechanism for measuring the semantic similarity between sentences.

Moreover, the effectiveness of each optimization strategy, i.e., LSF, SC, and KMA, is evaluated and shown in Fig. 2. The results indicate different characteristics of the techniques and furthermore, each of them has good performance. The superior of the integrated model demonstrates the complementarity of the three optimization strategies.

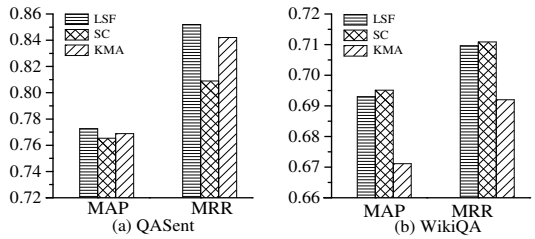

Figure 2: Performance comparison of the three optimization strategies on *QASent* and *WikiQA*

4. CONCLUSIONS

We have proposed an enhanced CNN model for answer selection. The intuitive idea is that the granular representations with more semantic features are deliberately designed and the effectiveness of the model has been experimentally evaluated on the benchmark datasets.

Acknowledgments: This research is supported by the National Natural Science Foundation of China: U1636116, 11431006, 61070089, Research Fund for International Young Scientists: 61650110510, Ministry of Education of Humanities and Social Science: 16YJC790123, Natural Science Foundation of Tianjin: 14JCYBJC15700, 15JCYBJC46600.

5. REFERENCES

[1] A. Severyn and A. Moschitti. Learning to rank short text pairs with convolutional deep neural networks. In *SIGIR*, pages 373–382, 2015.

[2] A. Severyn and A. Moschitti. Modeling relational information in question-answer pairs with convolutional neural networks. *arXiv:1604.01178*, 2016.

[3] M. Wang, N. A. Smith, and T. Mitamura. What is the jeopardy model? a quasi-synchronous grammar for qa. In *EMNLP*, pages 22–32, 2007.

[4] Z. Wang, H. Mi, and A. Ittycheriah. Sentence similarity learning by lexical decomposition and composition. In *COLING*, 2016.

[5] Y. Yang, W. tau Yih, and C. Meek. Wikiqa: A challenge dataset for open-domain question answering. In *EMNLP*, pages 2013–2018, 2015.

[6] W. Yin and H. Schütze. Convolutional neural network for paraphrase identification. In *NAACL*, 2015.

[7] L. Yu, K. M. Hermann, P. Blunsom, and S. Pulman. Deep learning for answer sentence selection. In *NIPS Deep Learning Workshop*, 2014.