Incorporating Context-Relevant Knowledge into Convolutional Neural Networks for Short Text Classification*

Jingyun Xu
South China University of Technology
Guangzhou,China
jingyun.x@qq.com

Yi Cai†
South China University of Technology
Guangzhou,China
ycai@scut.edu.cn

Abstract

Some text classification methods don't work well on short texts due to the data sparsity. What's more, they don't fully exploit context-relevant knowledge. In order to tackle these problems, we propose a neural network to incorporate context-relevant knowledge into a convolutional neural network for short text classification. Our model consists of two modules. The first module utilizes two layers to extract concept and context features respectively and then employs an attention layer to extract those context-relevant concepts. The second module utilizes a convolutional neural network to extract high-level features from the word and the context-relevant concept features. The experimental results on three datasets show that our proposed model outperforms the state-of-the-art models.

Introduction

Text classification is a task to classify a sentence into predefined categories, and it is widely applied in topic classification, sentiment analysis, and language inference. Most existing approaches use supervised machine learning approaches to establish classifiers, such as Support Vector Machine, Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN).

Although previous works achieve good performance, they don't work well on short texts due to the data sparsity. On the other hand, background knowledge plays an important role in natural language understanding. With the development of knowledge base construction, we can exploit these knowledge bases (KBs) to incorporate additional knowledge into neural networks. Moreover, the usefulness of knowledge features varies across contexts, as general KBs involve polysemy. For instance, the word 'Lincoln' can refer to a person or a car. Incorporating context-irrelevant knowledge into neural networks may mislead the classification result.

To address these problems, we propose a context-relevant concept recurrent convolutional neural network called CCR-CNN, which can incorporate context-relevant knowledge into a standard convolutional neural network. More specially, our model consists of two main modules. One is a lower sub-network: a *context-relevant concept representation module*. This module utilizes two layers to extract concept and context features respectively and then employs an attention layer to extract those context-relevant concepts. The other one is the upper sub-network: a *context-relevant concept word embedding based short text classification module*. This module combines word embedding and context-relevant concept embedding together called CCWE and feeds CCWE into a convolutional neural network.

Our Model

In this paper, we propose a context-relevant concept recurrent convolutional neural network called CCRCNN, which can capture context-relevant concepts features. Our model consists of two main modules, as details below.

Context-relevant Concept Representation Module

This module utilizes two layers to extract concept and context features respectively and then employs an attention layer to extract those context-relevant concepts.

Concept Representation Layer In Probase (Wu et al. 2012), knowledge is represented in the form of (word,concept) pair, which means that a word belongs to a concept. We utilize the words belong to a same concept to represent the concept. In other words, concept embedding is weighted average of word embedding:

$$v = \frac{\sum_{j=1}^{n} w_j^w e_j^w}{\sum_{j=1}^{n} w_j^w} \tag{1}$$

where e_j^w is the representation of word v_j^w, which is from the Google's pre-trained word vectors[1] and w_j^w represents the relevance of the word and the concept, which can be obtained from the Probase.

*This work was supported by the Fundamental Research Funds for the Central Universities, SCUT (No. 2017ZD048, D2182480), the Science and Technology Planning Project of Guangdong Province (No.2017B050506004), the Science and Technology Program of Guangzhou (No. 201704030076).

†Corresponding author: Yi Cai, ycai@scut.edu.cn

[1]https://code.google.com/archive/p/word2vec/

Context Representation Layer In the context representation layer, for a word x_t of the input at time t, we utilize a bidirectional GRU layer to get its two hidden states $\overrightarrow{h_t}$ and $\overleftarrow{h_t}$ and their concatenation $h_t = [\overrightarrow{h_t}; \overleftarrow{h_t}]$ provides a summary of global context around the input at time t.

Context-Relevant Concept Attention Layer Here we denote A to represent the concepts a word belong to in Probase, and only a part of A is relevant to a word's context. Therefore, given a word, we employ an context-aware concept attention layer to dynamically extract such concepts that are relevant to its context and aggregate the representations of those concepts to form a context-aware concept vector. Specifically, for a word x_t, with attention mechanism (Luong, Pham, and Manning 2015), its final concept representation is a weighted sum of its top 10 concepts:

$$c_t = \sum_{j=1}^{n} \alpha_{tj} v_{tj},\ with\ \alpha_{tj} = \frac{\exp(v_{tj} W_a h_t^T)}{\sum_{k=1}^{n} \exp(v_{tk} W_a h_t^T)} \qquad (2)$$

where v_{tj} is the representation for j^{th} concept for word x_t.

Context-relevant Concept Word Embedding based Short Text Classification Module

In this section, for a word at time t, we combine its representation x_t, and its concept representation c_t together called CCWE, and feed them into a convolutional neural network (Kim 2014) to transform the CCWE into a fixed size vector. Finally, we feed the output vector into a softmax layer to predict the class distribution. The training objective is to minimize the categorical cross-entropy loss of the predicted and true class distributions.

Experiments

Datasets

TREC. This dataset is a question dataset, which classify sentences into 6 question types, including person, location, numeric information, etc.
Movie Review (MR) This dataset contains 10662 sentences, each sentence is a positive or negative comment on movies.
AG. This dataset includes title and descriptions of AG's corpus of news, we only use the titles of each news in our experiment followed by (Wang et al. 2017).

Baseline methods

We compare our models with several state-of-the-art baseline models, including Bow+SVM (Chris Manning et al. 2012), CNN (Kim 2014), CharCNN (Zhang, Zhao, and LeCun 2015), CNN-non-static+UNI (Li et al. 2017), KPCNN (Wang et al. 2017).

Implementation details

In the experiments, we use 300-dimensional word2vec[2] vectors to initialize the word embedding. We use Adadelta to optimize the training process.

[2] https://code.google.com/archive/p/word2vec/

Experimental results

We use accuracy of the prediction as the evaluation metric, which is a stanford metric (Manning 2002). The experiment results are shown in Table 1. We can see that our model significantly outperforms state of the art methods. The basic CNN model can achieve good performance and our concept embedding can bring richer semantic information into word representation. In other words, concept information can be taken as a kind of prior knowledge to improve the performance of deep learning model in short text classification.

Table 1: Evalution results

Model	TREC	MR	AG
Bow + SVM	85.66	77.52	72.7
CNN	93.4	82.42	86.8
CharCNN	76	77.01	78.27
CNN-non-static+UNI	94.4	82.1	–
KPCNN	93.46	83.25	88.36
CCRCNN	**94.6**	**84.6**	**88.9**

Conclusion

In this paper, we propose a neural network CCRCNN, which firstly utilizes two layers to extracts concept and context features respectively and then employs an attention layer to extract those context-relevant concepts. Then concept features are incorporated into a convolutional neural network for short text classification. The experimental results on three datasets show that our proposed model outperforms state-of-the-art models.

References

Chris Manning, Sida Wang, T. G. A.; Moffat, A.; Webber, W.; and Zobel, J. 2012. Baselines and bigrams: Simple, good sentiment and topic classification using bag of features + linear classifiers. In *ACL*, 90–94.

Kim, Y. 2014. Convolutional neural networks for sentence classification. In *EMNLP*, 1746–1751.

Li, S.; Zhao, Z.; Liu, T.; Hu, R.; and Du, X. 2017. Initializing convolutional filters with semantic features for text classification. In *EMNLP*, 1884–1889.

Luong, T.; Pham, H.; and Manning, C. D. 2015. Effective approaches to attention-based neural machine translation. In *EMNLP*, 1412–1421.

Manning, C. D. 2002. Foundations of statistical natural language processing-table of contents. 4(1):80–81.

Wang, J.; Wang, Z.; Zhang, D.; and Yan, J. 2017. Combining knowledge with deep convolutional neural networks for short text classification. In *IJCAI*, 2915–2921.

Wu, W.; Li, H.; Wang, H.; and Zhu, K. Q. 2012. Probase: a probabilistic taxonomy for text understanding. In *SIGMOD Conference*, 481–492.

Zhang, X.; Zhao, J. J.; and LeCun, Y. 2015. Character-level convolutional networks for text classification. In *NIPS*, 649–657.