ATLANTIS PRESS
Advances in Intelligent Systems Research, volume 168
International Conference on Modeling, Analysis, Simulation Technologies and Applications (MASTA 2019)

A Method for Chinese Entity Relationship Extraction Based on Bi-GRU

Jian-qiong XIAO*, Zhi-yong ZHOU and Xing-xian LUO
Educational Information Technical Center, China West Normal University, China
*Corresponding author

Keywords: Bi-GRU, Regional list embedding(RLE), Hybrid neural network.

Abstract. In order to solve some defects of single deep neural network in Chinese entity Relationship Extraction task, a hybrid neural network entity relationship extraction model is designed and implemented in this paper. The model combines convolution network and bidirectional GRU model with a unified architecture, by defining varisized regional list embedding, it produces nonobjective feature representations of word vectors in distinction positions, and it has only Chinese character vectors and Chinese character word vectors, without position embedding. The laboratory findings show that our method is very effective on the Chinese corpus ACE2005 dataset about entities extraction task.

Introduction

Aiming at the problem that the traditional CNN model ignores the text context and leads to the lack of text semantics, a convolution layer improvement algorithm is proposed in this paper. By stripping the convolution layer from the CNN model, the convolution layer structure is improved, by defining a varisized sizes regional list embedding, and producing nonobjective feature representation of word vectors at imparity locations, which results in more accurate feature representation by fusing multiple local features, and position vectors aren't required.

Related Work

Many researchers have putted forward to a number of solid relational extraction methods based on deep neural networks.

Liu et al. [1] is first team who used convolution neural network to automatically learn sentence representation for relational classification tasks, the characteristics of lexical features, lexicality and so on are added to the model, their model get F1 value in the corpus ACE2005 dataset, and exceed the kernel function method 9%. Dong-xu Zhang et al. [2] used RNN to get varisized location feature by training corpus. Zhang [3] proposed using bidirectional long short-term memory network(Bi-LSTM) to build whole sentences extract model, this model used many features, include NLP tools and lexical resources, POS, NER and so on, of course, it achieved the state-of-the-art result.

In order to make full use of the competitive advantages of existing neural networks, Sun ziyang et al. [4] also used Bi- LSTM to model sentence dependency shortest path, and this model use out of CNN as input of LSTM to train the model. This method took full comprehensive advantage of Bi-LSTM model and the CNN model, because Bi-LSTM is good at capturing distance mutual relationship, while CNN can capture the local flat characteristics of sentences.

In view of above research, we propose a method for Chinese entity relationship extraction based on bidirectional Gated Recurrent Unit (RLE-BiGRU), it is a new means by defining a new vector (RLE). Experiments result is F1-score of 85.3% on the Chinese corpus ACE2005 dataset, which shows the validity of the work in this paper for the Chinese entity relations extraction.

Model

In this section we introduce our RLE-BiGRU model in detail. Figure 1 is the model photograph. Our model has six constituent parts:

(1) Input layer: input corpus, include sentence and Chinese character and Chinese vocabulary is to be tokenized;

(2) Embedding layer: map Chinese character and Chinese vocabulary within origin sentence into a low dim express;

(3) CNN: to get forward a single step local features by using express of step (2);

(4) Bi-GRU: to build higher level global sentences express and Chinese character and Chinese vocabulary local express within this sentences from step (3);

(5) Word-level attention layer: merge word-level features and character-level features from each time step to get their corresponding weights.

(6) Output layer: Get the probability of each entity category through Softmax.

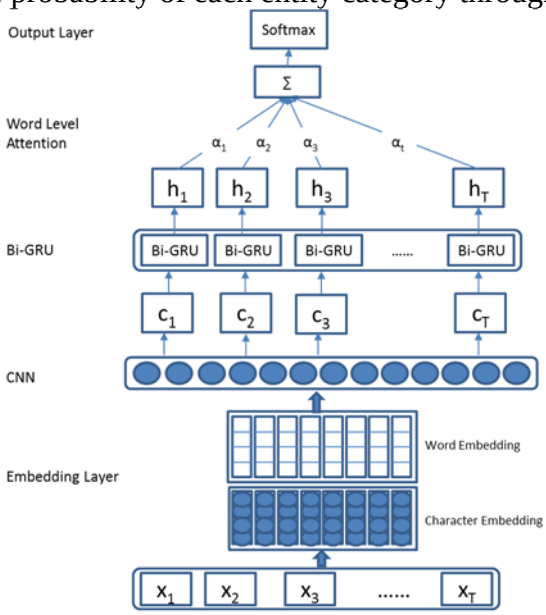

Figure 1. RLE-BiGRU model

Our model defines four networks, including two CNN, one Bi-GRU, and one Softmax. One of the CNN uses character-level features to extract Chinese characters and make the following improvements CNN: defining a regional list embedding RLE, original order word of sentence corresponding to the input statement is maintained, then through a one-filter convolution layer to extract the local feature representation of varisized locations. By processing three same dimension filters, they can produce multiple different abstract feature mapping P_{ij} at position i, and then perform convolution operations to produce a feature matrix similar to the position vector:

$$P = \begin{bmatrix} P_{11} & \cdots & P_{1n} \\ \vdots & \ddots & \vdots \\ P_{t1} & \cdots & P_{tn} \end{bmatrix}$$

Through this improvement, on the one hand, using multiple filters of different sizes to produce multiple local features on the word vector of position I, on the other hand, the newly generated abstract feature sequences still maintain the same contextual lexical logic relationship as the original text, which plays an important impact in improving relational classification effect. And the abstract feature sequence is used as the input of the next layer of GRU, which can realize to assemble convolution layer and Bi-GRU with a unified structure.

Embedding Layer

Our method has two embedding, namely Chinese character embedding and Chinese word (Chinese vocabulary) vectors. Chiu and Nichols [5] et al. used CNN to extract character-level features, and they achieved good goal in entity recognition. We use the means in [6]. First, we get Chinese character features by using our modified CNN to train, then use word2vec to pre-train Chinese word vector (original word vector, $\bar{V}_i^w$) from the corpus ACE2005, and then average Chinese character vectors and Chinese word vector as a new different word vector $\widetilde{V}_i^w$ of Chinese word, finally, $\bar{V}_i^w$

and $\widetilde{V}_i^w$ are put together to as final input embedding v_i^w of model.

Bi-GRU Layer

$$\begin{cases} z_t = \sigma(W_z \cdot [h_{t-1}, p_t] \\ r_t = \sigma(W_r \cdot [h_{t-1}, p_t] \\ \tilde{h}_t = \tanh(W_{\bar{h}} \cdot [r_t * h_{t-1}, p_t]) \\ h_t = (1 - z_t) * h_{t-1} + z_t * \tilde{h}_t \end{cases} \tag{1}$$

LSTM can overcome RNN's imperfection in long-distance dependence aspect, GRU (Gated Recurrent Unit) is a variant of LSTM, Due to GRU retains the LSTM merit its structure is simpler, its parameters are reduced and training rate are higher, so our model choose GRU, and we use bidirectional Bi-GRU structure. The calculation formula is shown in (1).Where P_t is output of CNN.

Attention Layer

Our model has two types attention, which used to acquire the importance degree of words and sentences in documents respectively. One is Chinese character-level attention which is embedded in BiGRU, model for each input sentence to do training, add the word level attention, and capture the keyword in a sentence. Another is Chinese vocabulary level attention for each category input sentence to do common training, add the word level attention model, capture the entity front and back pivotal language environment.

$$\begin{cases} \vec{h}_{it} = \overrightarrow{GRU}(c_{it}), t \in [1, T] \\ \overleftarrow{h}_{it} = \overleftarrow{GRU}(c_{it}), t \in [T, 1] \\ u_{it} = \tanh(W_w \cdot [\vec{h}_{it}, \overleftarrow{h}_{it}] + b_w) \\ \theta_{it} = \dfrac{\exp(u_{it}^T u_w)}{\sum_t \exp(u_{it}^T u_w)} \\ s_i = \sum_t \theta_{it} h_{it} \end{cases} \tag{2}$$

A word sequence and sentence list are encoded forward direction and reverse direction by GRU and attention was get, they are represented as (2)

$$\mathrm{U} = \tanh(\mathrm{H}) \tag{3}$$

$$\alpha = \mathrm{softmax}(w^T U) \tag{4}$$

$$\mathrm{r} = \mathrm{H} \cdot \alpha^T \tag{5}$$

Experiments

Dataset and Experimental Setup

Our experiment use Chinese corpus ACE2005 Chinese entity extraction task dataset. It has 6 categories of relationships and 9255 entity relationship. 4/5 of them were extracted as training data

according to the category scale stochastic, and rest dataset 1/5 as test data. The verification set from the training set random sampling obtained.

In order to assess the consequence of our model in this paper, four existing methods are selected as comparisons. Table 1 is the comparative result. The comparison model is as follows: RNN model proposed by Zhang and Wang, the BLSTM model proposed by Zhang and Zheng, the Att-BLSTM model designed by Peng Zhou et al., and the model used by Sun Zi-yang et al.

Our model achieved good results, and its F1 value reached 85.3%, indicating effectiveness of this method. Our model is simpler to above-mentioned model for comparisons, without lexical resources or NLP systems. The different sizes regional list embedding can accurately produce accuracy feature representations of word vectors in different positions, thus eliminating the need for additional position vectors.

Conclusion

In this paper, we proposed a different neural network Chinese entities extract model, named RLE-BiGRU. The experimental results show that our method can exceed most of the existing methods, with only Chinese character vectors and Chinese word vectors and regional list embedding, no any additional features, could get high-level global features.

Table 1. Comparison results

Model	Feature Set	F1(%)
RNN(Zhang and Wang,2015)	Word vectors (dim=300) + Position vectors	82.8
BLSTM(Zhang et al.,2015)	Word vectors(dim=200) + Position features +other speech characteristics	84.3
Att-BLSTM(Peng Zhou)	Word vectors(dim=100)	84.0
Sun Zi-yang	Bag-of-POS +SDP	82.7
RLE-BiGRU	word vectors(dim=50) +character embedding+ RLE	85.3

Acknowledgments

This research was supported by the Innovation Team of China West Normal University (No.CXTD2017-6) and Excellence Fund of China West Normal University (No.17Y183).

References

[1] Liu C Y,Sun W B,Chao WH,et al.Convolution neural network for relation extraction[C]// International Conference on Advanced Data Mining and Applications[J].2013:231-242.

[2] Zhang D,Wang D.Relation Classification via Recurrent Neural Network[EB/OL]. https://arxiv.org/pdf/1508.01006. pdf,2015-04-05.

[3] Zhang S,Zheng D,Hu X,etal.Bidirectional Long Short-Term Memory Networks for Relation Classification[C]//PACLIC, 2015

[4] SUN Ziyang, GU Junzhong, YANG Jing,et al. Chinese Entity Relation Extraction Method Based on Deep Learning[EB/OL]. http://www.ecice06.com/CN/abstract/abstract28113. shtml,2017.

[5] Chiu J P C,Nichols E.Named Entity Recognition with Bidirectional LSTM-CNNs[J].2015, arXiv preprint arXiv: 1511.08308.

[6] Daojian Zeng, Kang Liu, Siwei Lai, Guangyou Zhou, and Jun Zhao.2014.Relation Classification via Convolutional Deep Neural Network. In Proceedings of the 25th International Conference on Computational Linguistics (COLING), pages2335–2344.