


MULTI-CHANNEL CONVOLUTIONAL NEURAL NETWORK FOR TARGETED SENTIMENT CLASSIFICATION

TING YUAN¹, HAI-HUI LI^{1,2}, HONG-YA ZHAO², QIAN-HUA CAI¹, HAN LIU³, XIAO-HUI HU¹

¹ South China Normal University, Guangzhou 510006, China

² Shenzhen Polytechnic, Shenzhen 518055, China

³ Cardiff University, Cardiff CF24 3AA, United Kingdom

E-MAIL: yuantingyt@m.scnu.edu.cn, lihaihuisncnu@m.scnu.edu.cn, hy.zhao@szpt.edu.cn, caiqianhua@m.scnu.edu.cn, liuh48@cardiff.ac.uk, huxh@scnu.edu.cn

Abstract:

In recent years, targeted sentiment analysis has received great attention as a fine-grained sentiment analysis. Determining the sentiment polarity of a specific target in a sentence is the main task. This paper proposes a multi-channel convolutional neural network (MCL-CNN) for targeted sentiment classification. Our approach can not only parallelize over the words of a sentence, but also extract local features effectively. Contexts and targets can be more comprehensively utilized by using part-of-speech information, semantic information and interactive information, so that diverse features can be obtained. Finally, experimental results on the SemEval 2014 dataset demonstrate the effectiveness of this method.

Keywords:

Attention mechanism; Targeted sentiment analysis; Convolutional neural networks; Multi-channel

1. Introduction

With the rise of social networks and e-commerce platforms, more and more people have been interested in posting comments through the Internet. Sentiment analysis is one of the important research areas in Natural Language Processing (NLP), and its purpose is to determine the reviewers' attitude against comments [1]. The traditional sentiment analysis task is mainly aimed to obtain the overall sentiment of the sentence. When there are multiple targets in the sentence, the sentiment of each target is ignored. Therefore, methods for achieving the targeted sentiment analysis have been proposed [2]. For example, given the sentence "The appetizers is ok, but the service is slow and the environment is bad." there are three targets: "appetizers", "service" and "environment". Their sentiments are positive, negative and negative, respectively. Targeted sentiment analysis shows its advantage and necessity when a sentence contains multiple targets.

In recent years, deep neural networks have been widely used in targeted sentiment analysis and achieved great results.

Liang proposed a multi-attention convolutional neural networks (MATT-CNN) [6]. It can capture deeper level sentiment information and distinguish sentiment polarity of different targets explicitly through a multi-attention mechanism leading to great results. It has better enabled the task of targeted sentiment analysis, but there are still some challenges as follows. Firstly, contexts or targets usually contain more than one word, and different words play different roles in sentiment analysis. MATT-CNN cannot make full use of the differences among words. Secondly, the information used is too monotonous to extract diverse features. Thirdly, it is impossible to process different kinds of information at the same time, resulting in the time wasting. In this paper, a multi-channel convolution neural network approach is proposed to address the above problems.

The main contributions include: Firstly, a part of speech (POS) based attention mechanism is proposed. The impact of the target POS on the context and the target can be identified through learning from the parameter matrix and a similar impact can result from the use of the context POS. Secondly, an Euclidean distance based attention mechanism is proposed, which makes full use of the semantic information and interacts with the target and the context as semantically as possible. Thirdly, more features and local information can be extracted by making full use of the target and the context information by interactive learning between the target and the context, and by using multi-channel CNN.

Model's verification is carried out using the dataset of laptop and restaurant in SemEval 2014 Task4 [7]. The results show that this method could achieve better classification in the task of targeted sentiment analysis.

2. Related Work

In recent years, the rise of deep neural networks has promoted the development of targeted sentiment analysis, which can combine the textual features [8]. [3] separated the

sentences into two parts from the target and modeled with two independent LSTM models. It can make full use of the semantic information of the target and capture the relationship between the target and contexts. However, this model did not take into account the differences among words. [4] used the LSTM model after stitching the target and each word in the contexts to obtain the hidden states, then applied attention network to determine the influence of a specific target on the sentence, which led to increasing the weight of important words through the attention networks, but cannot extract local information. Therefore, [6] used MATT-CNN captured deeper level sentiment information and distinguished sentiment polarity of different aspects explicitly through a multi-attention mechanism without using any external parsing results. It can get better performance than the ones obtained using traditional CNN, attention-based CNN and attention-based LSTM. However, this method cannot achieve extraction of features from multiple information channels at the same time.

Therefore, we have proposed a multi-channel CNN based sentiment analysis method with POS information, semantic information and interactive information to achieve more effective targeted sentiment analysis, which adds information from different perspectives and obtains diverse features, thus effectively distinguishing the effects of different words on sentiment analysis.

3. Model

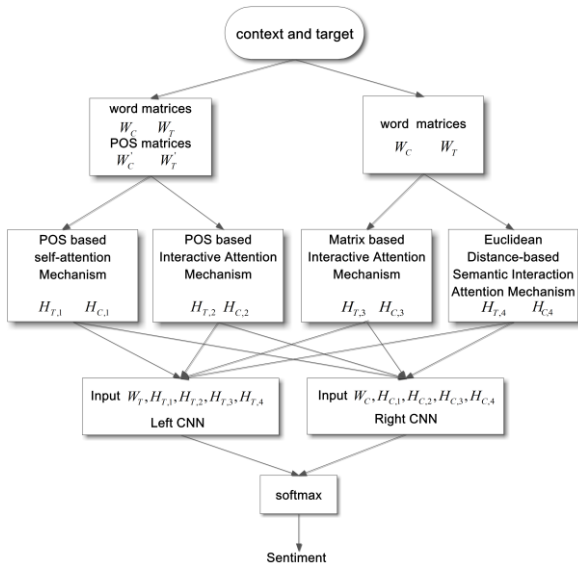


Fig. 1 structure of MCL-CNN

In this paper, a multi-channel convolution neural network approach is proposed, which can make full use of the target and context information and extract diverse features, so

as to make targeted sentiment analysis more effective. The structure of the multi-channel CNN is shown in Fig. 1.

3.1. Based on POS Attention Mechanism

In general, POS is very capable to play a very important role in judging sentiments. For example, the adjectives "good" and "bad" represent two opposite sentiments, respectively. It is obvious that POS is particularly important. Therefore, this paper proposes a CNN approach based on POS attention mechanism, which trains POS vectors for each part of speech, and then learns the corresponding weights of words through POS. Next, it introduces the steps involved in implementing the POS based attention mechanism.

Firstly, the word vector is mapped, and the target matrix $W_T \in \mathbb{R}^{m \times d}$ and the context matrix $W_C \in \mathbb{R}^{n \times d}$ are obtained, where m and n are the number of words in the target and the number of words in the context, respectively, and d is the dimension of word embedding. The POS vectors for each word are trained by using word2vec, and then mapping the target POS matrix $W_T' \in \mathbb{R}^{m \times d'}$ and the context POS matrix $W_C' \in \mathbb{R}^{n \times d'}$, d' is POS vector dimension.

Secondly, the weight of each word is derived through POS based attention mechanism. Targeted sentiment analysis needs both target and context information, so the POS of the target has an impact on both the target and the context. The POS of the context has a similar impact.

3.1.1. POS based self-attention Mechanism

Firstly, through using the POS based self-attention mechanism of the target matrix and the target POS matrix, the weight matrix of the attention mechanism of the target is obtained. The formula is as follows:

$$ATT_{t,1} = \tanh(W_T' \cdot W_{t,1} \cdot W_T^T) \quad (1)$$

$W_{t,1} \in \mathbb{R}^{d' \times d}$ is parameters matrix, W_T^T is the transpose matrix of W_T .

All elements of $ATT_{t,1}$ matrix are normalized, and an attention weight matrix $A_{t,1}$ is obtained. such that, the results of using target POS based self-attention mechanism $H_{T,1}$ are obtained according to the formula as follows:

$$H_{T,1} = A_{t,1} \cdot W_T \quad (2)$$

$H_{T,1} \in \mathbb{R}^{m \times d}$. The i th row vector of the $H_{T,1}$ matrix represents the weighted result of the i th word in target POS for all target vectors. The formula is as follows:

$$h_{T,1}^i = \sum_{j=1}^m \alpha_{i,1}^{i,j} w_T^j \quad (3)$$

w_T^j is the j th word vector in W_T and $\alpha_{i,1}^{i,j}$ is the j th weight in $A_{i,1}$.

Through using POS based self-attention mechanism, the target POS information is included in the target vector.

The context information is obtained in a way similar to the one in which the target information is obtained, through using the formula as follows:

$$\begin{aligned} ATT_{c,1} &= \tanh(W_C^i \cdot W_{c,1} \cdot W_C^T) \\ H_{c,1} &= B_{i,1} \cdot W_C \end{aligned} \quad (4)$$

$W_{c,1} \in \mathbb{R}^{d \times d}$, W_C^T is the transpose matrix of W_C , $B_{c,1}$ is the weight matrix which is obtained through normalization of $ATT_{c,1} \in \mathbb{R}^{n \times n}$.

In this way, the context POS information is included in the context vector.

3.1.2. POS based Interactive Attention Mechanism

The difference between the POS based interactive attention mechanism and the POS based self-attention mechanism is on the interactive ways. Target POS has an influence on the context vectors, according to the formula as follows:

$$ATT_{i,2} = \tanh(W_T^i \cdot W_{i,2} \cdot W_C^T) \quad (5)$$

$W_{i,2} \in \mathbb{R}^{d \times d}$, W_C^T is the transpose matrix of W_C , $ATT_{i,2} \in \mathbb{R}^{m \times n}$. All elements of $ATT_{i,2}$ matrix are normalized, and an attention weight matrix $A_{i,2}$ is obtained, such that, the results of using the target POS based interactive attention mechanism is obtained using the formula is as follows:

$$H_{T,2} = A_{i,2} \cdot W_C \quad (6)$$

Similarly, context POS has an impact on the target vectors, according to the formula as follows:

$$\begin{aligned} ATT_{c,2} &= \tanh(W_C^i \cdot W_{c,2} \cdot W_T^T) \\ H_{c,2} &= B_{c,2} \cdot W_T \end{aligned} \quad (7)$$

$W_{c,2} \in \mathbb{R}^{d \times d}$, $H_{T,2} \in \mathbb{R}^{m \times d}$, $H_{c,2} \in \mathbb{R}^{n \times d}$, W_T^T is the transpose matrix of W_T , $ATT_{c,2} \in \mathbb{R}^{n \times m}$. $B_{c,2}$ is a weight matrix.

POS based interactive attention mechanism makes the target POS information fully act on the context, so that the $H_{T,2}$ matrix contains both the context information and the target POS information and the $H_{c,2}$ does as well.

3.2. Interactive attention mechanism

How to make full use of the above two types of information is particularly important in targeted sentiment analysis. In this section, we applies two interactive attention mechanisms, namely, the matrix-based interactive attention mechanism and the proposed interactive attention mechanism based on Euclidean distance, to design the multi-channel convolutional neural network model.

3.2.1. Matrix-based Interactive attention mechanism

We use the matrix-based interactive attention proposed by [11], which can get great interactive information according to the formula as follows:

$$TC = \tanh(W_T \cdot W_{ic} \cdot W_C^T) \quad (8)$$

$W_{ic} \in \mathbb{R}^{d \times d}$ represents attention parameters matrix, W_C^T is the transpose matrix of W_C , $TC \in \mathbb{R}^{m \times n}$.

We obtain interactive weight coefficient of target to context T by row normalization of TC , such that the results of using interactive attention mechanism $H_{T,3} \in \mathbb{R}^{m \times d}$ are obtained according to the formula as follows:

$$H_{T,3} = T \cdot W_C \quad (9)$$

After interactive learning, matrix $H_{T,3}$ is obtained, which results in context-specific target information.

The interactive attention matrix TC is transposed as TC^T and we can get C by row normalization of TC^T , such that the results of interactive attention mechanism $H_{C,3} \in \mathbb{R}^{n \times d}$ are obtained according to the formula as follows:

$$H_{C,3} = C \cdot W_T \quad (10)$$

After interactive learning, matrix $H_{C,3}$ is obtained, which results in target-specific context information.

3.2.2. Euclidean Distance-based Semantic Interaction Attention Mechanism

In the process of sentence matching, [5] points out that it's flawed to model a single sentence. Therefore, Euclidean distance based interaction between two sentences is proposed. Traditional interactive attention gets the weight coefficients by multiplying two different word vectors interactive learning, which cannot use semantic distance information. Therefore, this paper proposes a CNN approach based on Euclidean distance which links between targets and contexts as:

$$d_{i,j} = \frac{1}{1 + \text{dist}(w_T^i - w_C^j)} \quad (11)$$

w_T^i is i th word vector of W_T , w_C^j is j th word vector of W_C .

$d_{i,j}$ is an element in $D \in \mathbb{R}^{m \times n}$, which shows the interaction between the i th word in the target and the j th word in the context. The distance means that the two words with high similarity have greater impacts on each other, and the weight of the attention mechanism is correspondingly larger.

Then, for getting target weight $a \in \mathbb{R}^m$ based on Euclidean distance, the formula is shown as follows:

$$a = D \cdot u_i \quad (12)$$

$u_i \in \mathbb{R}^n$ is a parameter vector. It means that all contexts based on Euclidean distance information assign a weight to each target word vector. Then, the weight coefficient is multiplied with the corresponding target word vector to get target matrix $H_{T,4}$ containing semantic information, according to formula as follows:

$$H_{T,4} = a \circ W_T = (a_1 \cdot w_T^1 \quad a_2 \cdot w_T^2 \quad \dots \quad a_m \cdot w_T^m) \quad (13)$$

Where \circ represents constants multiplied by vectors, a_i is i th element in a , w_T^i is i th target vector of W_T , $i \in [1, m]$, after using the European distance based interactive attention mechanism. $H_{T,4}$ includes the target information, context information and semantic information.

Then, context weight b based on Euclidean distance is obtained according to the formula as follows:

$$b = D^T \cdot u_c \quad (14)$$

$u_c \in \mathbb{R}^m$ represents attention parameters vector, $b \in \mathbb{R}^n$, which means that all target vector based on Euclidean distance information assigns a weight to each context word vector, such that the context matrix $H_{C,4}$ is obtained. using the formula as follows:

$$H_{C,4} = b \circ W_C = (b_1 \cdot w_C^1 \quad b_2 \cdot w_C^2 \quad \dots \quad b_n \cdot w_C^n) \quad (15)$$

b_j is i th element in b , w_C^j is j th context vector of W_C , $j \in [1, n]$, after using Euclidean distance based interactive attention mechanism, $H_{C,4}$ includes the context information, target information and semantic information.

3.3. Multichannel CNN

Convolutional neural networks can be used leading to effective reduction of the computational complexity by sharing weights among different convolution kernels, and thus it can obtain local features through local perception. A convolution neural network is usually composed of several convolution layers, pooling layers and fully connected layers.

At present, multi-channel CNN is widely used in the field of image processing, and the information of each

channel is different. In this paper, multi-channel CNN is used to achieve more effective targeted sentiment analysis. The multi-channel CNN model example is shown in Fig. 2.

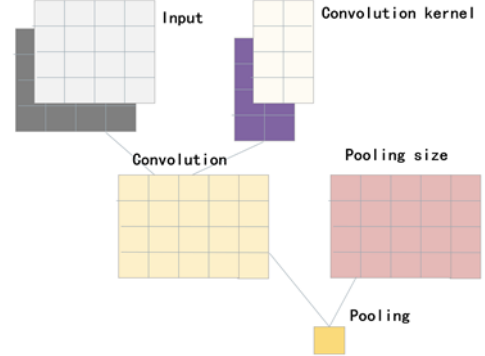


Fig. 2 multi-channel CNN model

By splicing matrixes resulting from POS attention and interactive attention mechanisms, respectively, in the third dimension, the inputs of two multi-channel CNNs can be obtained. $W_C, H_{C,1}, H_{C,2}, H_{C,3}, H_{C,4}$ splicing results in the input of one CNN, and $W_T, H_{T,1}, H_{T,2}, H_{T,3}, H_{T,4}$ splicing results in the input of the other CNN.

The specific parameters of the two multi-channel CNNs in the model are shown in Table 1.

Table 1 CNN specific parameters

Parameter	Left CNN	Right CNN
Input	$(d, m, 5)$	$(d, n, 5)$
Convolution	$(d, w, 5)$	$(d, w, 5)$
Convolution result	$(d, m + w - 1)$	$(d, n + w - 1)$
Pooling	$(d, m + w - 1)$	$(d, n + w - 1)$

The third dimension of input matrix "5" represents five channels, (d, m) is the size of target matrix, (d, n) is the size of context matrix, d is the dimensionality of word vectors, m and n are the number of words in the target and the number of words in the context, respectively.

The operation of the convolution part is to fill the left and right sides of the original input matrix with 0 matrices of $(d, m + w - 1)$ size, without processing the upper and lower sides. It can ensure the utilization of the target and the context. w is a hyper parameter, and s represents convolution kernels. Then, the whole result matrix is pooled. Finally, the final target representation $t_{all} \in \mathbb{R}^s$ and the final context representation $c_{all} \in \mathbb{R}^s$ are obtained.

Splicing t_{all} and c_{all} obtains r , which is sent into the softmax layer. Finally, the targeted sentiment analysis results

are obtained according to the formulas below

$$\begin{aligned}
 x &= \tanh(r \cdot W_r + b_r) \\
 y_i &= \frac{\exp(x_i)}{\sum_{i=1}^C \exp(x_i)}
 \end{aligned} \tag{16}$$

$W_r \in \mathbb{R}^{2s \times C}$ is parameter weight, $b_r \in \mathbb{R}^C$ is bias.

4. Model training

We use cross entropy with L_2 regularization as the loss function. The formula is as follows:

$$loss = -\sum_{i=1}^C y_i \log g_i + \lambda_r \left(\sum_{\theta \in \Theta} \theta^2 \right) \tag{17}$$

Where $y_i \in \mathbb{R}^C$ denotes the truth sentiment polarity, $g_i \in \mathbb{R}^C$ is the predicted sentiment analysis result. C represents the total number of sentiment classification and λ_r is the parameter of L_2 regularization.

Then, the back propagation method is used to calculate the gradients and update all the parameters Θ :

$$\Theta = \Theta - \lambda_r \frac{\partial J(\Theta)}{\partial \Theta} \tag{18}$$

5. Experiments

5.1. Dataset

We use the data set provided in SemEval 2014 Task4 to validate the effectiveness the proposed MCL-CNN model. This dataset includes reviews for laptops and restaurants, and they are both three-class classification problems. The training and test reviews are shown in Table 2.

The performance of MCL-CNN is evaluated in terms of accuracy, according to the formula as follows:

$$Acc = \frac{T}{N} \tag{19}$$

where T is the number of correctly predicted test instances, N is the total number of test instances.

Table 2 Statistics of SemEval 2014 Dataset

Dataset	Positive	Neutral	Negative
Lap-Train	994	464	870
Lap-Test	341	169	128
Res-Train	2164	637	807
Res-Test	728	196	196

5.2. Hyper parameters setting

We use the pre-trained GloVe model for transforming targets and contexts into word vectors of 300 dimensions. The initial values of the attention parameters are generated by uniform distribution $U(-0.01, 0.01)$ and all the initial biases are set to 0. Meanwhile, the learning rate is set to 0.01 and the dropout is set to 0.5 to avoid over-fitting. The number of convolution kernels is 16, and the size is (300, 4, 2).

5.3. Results

In order to prove the validity of the proposed MCL-CNN model, we compares the model with several other ones. The classification results are shown in Table 3.

(1) LSTM: Through single-level LSTM modeling, targeted sentiment analysis is carried out.

(2) TD-LSTM: Separating the sentences into two parts from the target and modeling with two independent LSTM models, splicing output to get target and context information [3].

(3) AT-LSTM: Firstly, LSTM is trained, and LSTM's hidden layer output is spliced with the target. Then, the final feature representation is obtained by using attention mechanism [4].

(4) CNN: The simplest convolutional neural network is used to model and analyze the sentiment of a specific target [9].

(5) ATT-CNN: Multi-layer CNN is used for feature extraction and attention mechanism is introduced [10].

(6) CATT-CNN: On the basis of convolution neural networks, a variety of attention mechanisms are introduced to targeted sentiment analysis [6].

Table 3 Comparison of experimental results

Methods	Restaurant	Laptop
LSTM	0.743	0.665
TD-LSTM	0.756	0.681
AE-LSTM	0.762	0.689
CNN	0.663	0.583
ATT-CNN	0.759	0.654
CATT-CNN	0.792	0.671
MCL-CNN	0.794	0.721

As shown in Table 3, the original LSTM and the original CNN are the worst performing ones on the two datasets in all experiments. However, through introducing the attention mechanism, the effect is obviously improved. The use of TD-LSTM and AE-LSTM, have lead to improved accuracy rates by about 1.3%-1.9% and 1.6%-2.4%, respectively, compared with the use of the original LSTM. The use of

ATT-CNN and CATT-CNN have lead to improved accuracy rates by about 9.6%-12.9% and 7.1%-8.8%, respectively, compared with the use of the original CNN.

It can be seen from the experimental results that the MCL-CNN approach proposed in this paper achieves the best performance in the current experiments, especially on the laptop dataset, showing great success.

6. Conclusion

The information is limited in the setting of traditional methods of targeted sentiment analysis. Therefore, POS information is added through the adoption of POS attention. Through involving interaction, the target contains the context information, and the context also contains the target information. Moreover, the hierarchical semantic relationship of the Euclidean distance attention mechanism is used to add the semantic information, so as to make the information fully available. Multichannel CNN is selected and the obtained information is sent into two CNNs to extract diverse features. The experimental results show that the proposed approach is effective in undertaking targeted sentiment analysis tasks and achieves better results than traditional methods.

There are more extensive works that can be done in the future on the task of targeted sentiment analysis. One of the focuses in our future research is to obtain more helpful representations of sentiment analysis through the interactive learning of contexts and target. We will consider how to make full use of the advantages of multiple in-depth learning.

References

- [1] X.Glorot, A.Bordes, and Y.Bengio, "Domain adaptation for large-scale sentiment classification: A deep learning approach," in *Proceedings of the 28th international conference on machine learning*, Bellevue, Washington, USA, 28 June-2 July 2011, pp.513-520.
- [2] S.Chen, C.Peng, L.Cai, and L.Guo, "A Deep Network Model for Specific Target Sentiment Analysis," *Computer Engineering*, vol.45, no.3, pp.286-292, 2019.
- [3] D.Tang, B.Qin, X.Feng, and T.Liu, "Effective LSTMs for Target-Dependent Sentiment Classification," in *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, Osaka, Japan, 13-16 December 2016, pp.3298-3307.
- [4] Y.Wang, M.Huang, and L.Zhao. "Attention-based LSTM for aspect-level sentiment classification," in *Proceedings of the 2016 conference on empirical methods in natural language processing*, Austin, Texas, USA, 1-5 November 2016, pp.606-615.
- [5] W.Yin, H.Schütze, B.Xiang, and B.Zhou, "Abcnn: Attention-based convolutional neural network for modeling sentence pairs," *Transactions of the Association for Computational Linguistics*, vol.4, pp. 259-272, 2016.
- [6] B.Liang, Q.Liu, J.Xu, Q.Zhou, and P.Zhang "Aspect-Based Sentiment Analysis Based on Multi-Attention CNN," *Journal of Computer Research and Development*, vol.54, no. 8, pp.1724-1735, 2017.
- [7] M.Pontiki, D.Galanis, H.Papageorgiou, I.Androustopoulos, S.Manandhar, A. S.Mohammad,,and V.Hoste, "Semeval-2016 task 5: Aspect based sentiment analysis," in *Proceedings of the 10th international workshop on semantic evaluation*, San Diego, California, 16-17 June 2016, pp.19-30.
- [8] S.Zheng and R.Xia, "Left-Center-Right Separated Neural Network for Aspect-based Sentiment Analysis with Rotatory Attention," *arXiv preprint arXiv:1802.00892*, 2018.
- [9] Y.Kim, "Convolutional neural networks for sentence classification," in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing*, Doha, Qatar, 25-29 October 2014, pp. 1746-1751.
- [10] L.Wang, Z.Cao, G. Melo, and Z.Liu, "Relation Classification via Multi-Level Attention CNNs," in *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. vol.1, Berlin, Germany, 7-12 August 2016, pp.1298-1307.
- [11] H.Li, T.Yuan, H.Wu, Y.Xue, and X.Hu, "Granular computing-based multi-level interactive attention networks for targeted sentiment analysis," *Granular Computing*, pp.1-9, 2019.