Research on Sentiment Analysis of Chinese E-Commerce Comments Based on Deep Learning

Hanchen Zhou, Gang Wu
Shanghai Jiao Tong University, Dongchuan Rd., Shanghai, China
zhc@sjtu.edu.cn

Abstract. The analysis of emotional polarity of Chinese phrases in complex contexts has always been a problem in the field of machine learning. The emergence of Artificial Neural Networks has opened a window to solve this problem. Sentiment analysis has also become a hotspot in the research. This paper aims to solve this problem by introducing Attention mechanism to the Temporal Convolutional Nets (TCN) model. This paper chooses the Chinese commodity comment phrase on the E-commerce platform as the research object, to automatically analyse and judge the polarity (positive and negative) of each comment phrase. This paper optimized the design and constructed of the TCN Attention model, which made the TCN Attention model show superior performance in the area of sentiment analysis, and improved significantly compared with other models.

1. Introduction

With the development of the Internet and the rise of E-commerce, people's shopping methods have undergone great changes. With the popularity and rapid growth of online shopping, the commodity comment information on the E-commerce platform has grown geometrically. It is obviously unrealistic to rely on manual statistics and analysis of these information. Therefore, E-commerce generally needs to design a special evaluation management system to analyse these comment to explore the potential value. At present, most evaluation management systems require users to comment on purchased products, and also allow users to rate or classify products according to their preferences (such as praise, average, and bad reviews). Due to the inaccuracy of computer classification of emotions, we can only rely on ratings or classifications instead of product evaluation results, which cannot represent the realistic and accurate evaluation of the user.

If the user only need to input the product review, and the computer help to automatically classify the input comments, named emotional polarity analysis [1], can be solved. This problem is a field of Natural Language Processing (NLP), which classify the input language as positive or negative. Due to the complexity of human language, traditional algorithms cannot perform effectively in sentiment analysis. Until the development of Artificial Neural Network model in the field of deep learning [2] in recent years, it has brought breakthrough progress to accurate sentiment analysis. The idea of Artificial Neural Network has been proposed in the last century. The basic principle is to let the computer simulate the operating mechanism of the human brain. With the rapid development of computer processing capabilities, deep neural network algorithms have made great progress. In the field of NLP, models based on Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN) have been proposed to solve problems in translation, question and answer, and sentiment analysis.

In the scenario of Chinese comment on E-commerce platform, we proposed a neural network model for Chinese sentiment analysis - TCN Attention, and compare it with LSTM [3], Bi-directional LSTM
[4], LSTM Attention [5] and TCN [6] model on the same data set. The results show that the TCN Attention are better than others.

2. Related Work

This paper applies Artificial Neural Networks technology to the emotional analysis of E-commerce platform commodity comments, aiming at mining the core value in commodity comment information, correctly guiding consumers to purchase goods, and making E-commerce play an active role in people's work and life.

Using artificial neural networks for natural language processing, Hinton first proposed using low-dimensional dense vectors to represent text [7], which is called a distributed representation. Distributed representations can effectively express the similarity between texts. Subsequently, Mikolov et al. proposed the Word2Vec tool [8], using the CBOW model and Skip-gram model to make the word to vector transformation more precise and effective.

At present, in the field of using neural networks to analyze sentiment, RNN has always performed well in processing time series. Liu et al. proposed to apply the LSTM model and the Bi-directional LSTM model to text classification tasks [4]. Yang et al. introduced the hierarchical attention networks into the text classification [9], and these models can be applied to sentiment analysis. However, the CNN has not had a good solution in processing time series until Bai et al. proposed the Temporal Convolutional Nets(TCN) structure [6], and proved that the model has better performance than RNN. These models have achieved good results in English sentiment analysis. But on Chinese data sets, these models perform not satisfactory. Therefore, this paper attempts to solve the problem of Chinese sentiment analysis through innovative thinking.

3. TCN Attention Based Sentiment Analysis Model Structure

The essence of the sentiment analysis model is a classifier. The input is segmented corpus, and the output is the sentiment category corresponding to the corpus. The model framework designed for sentiment analysis is shown in Fig.1. The model is a 5-layer system consisting of an input layer, an output layer and three hidden layers.

![Fig.1 The structure of sentiment analysis model](image)

**Embedding** layer converts index into a distributed word vector within the model, which converts a positive integer into a vector with fixed dimension. This layer accepts $1 \times T$ dimensional vector and outputs the matrix of $V \times T$. While initializing, it directly loads the projection layer weight matrix $W_{VN}$ of the CBOW model, and operate the index with follow steps: Generate a $N \times 1$ dimensional unit column vector $\vec{x}^T$ with the index$^{	ext{th}}$ dimension set to 1; Dot product $W_{VN}$ with $\vec{x}^T$. The result is shown in the form (1)
\[
\begin{bmatrix}
  w_{11} & w_{12} & \cdots & w_{1N} \\
  w_{21} & w_{22} & \cdots & w_{2N} \\
  \vdots & \vdots & \ddots & \vdots \\
  w_{V1} & w_{V2} & \cdots & w_{VN}
\end{bmatrix}
\begin{bmatrix}
  0 \\
  \vdots \\
  1 \\
  \vdots \\
  0 
\end{bmatrix}
= 
\begin{bmatrix}
  w_{1i} \\
  w_{2i} \\
  \vdots \\
  w_{Vi} 
\end{bmatrix}
\] (1)

In the formula, \( i \) represent index. The essence of the above formula is a process of looking up the table, extracting the vector of the index column, and each column of the matrix \( W_{V \times N} \) represents distributed representation of the corresponding vocabulary. Embedding layer make the Word2Vec process integrated into the sentiment analysis model.

The initialization parameters used by the Embedding layer is the weight of the CBOW model's projection layer. In fact, the Embedding layer can still be used for sentiment analysis training by initializing parameters with random numbers. But theoretically, a model that uses the CBOW model projection layer for Embedding layer initialization will perform better than the model use a random number for Embedding layer's initialization. This is due to the use of migration learning as a pre-training method. Migration learning refers to the knowledge learned in one environment being used in another to improve the generalization performance of the model. CBOW model is a linguistic probabilistic model, which is more focused on the vocabulary feature extraction. Transferring the training results from the CBOW model to the sentiment analysis model, allows the sentiment analysis model to converge faster and perform better.

TCN Attention layer is the core of sentiment analysis. This layer is used for time series' feature extraction, connecting with Embedding layer, accepting \( V \times T \) dimensional matrix as input and \( \vec{m} \) as output, the detailed structure of this layer will be given at next section.

Dense layer is used to classify the emotional features of the TCN Attention layer output. Dense layer is also called fully connected layer, which is the most basic neural network layer structure, consisting of several perceptrons connected in parallel. Dense layer accepts the output of TCN Attention layer as input, and obtain the output \( \vec{o} \). Let \( M \) denote the dimensions of the input feature vector \( \vec{m} \), \( O \) denote the dimension of the output vector \( \vec{o} \), and the operation of the Dense layer on the data can be used as form (2) illustrated.

\[
\sigma \left( \begin{bmatrix}
  w_{11} & \cdots & w_{1M} \\
  \vdots & \ddots & \vdots \\
  w_{O1} & \cdots & w_{OM}
\end{bmatrix}
\begin{bmatrix}
  m_1 \\
  \vdots \\
  m_M
\end{bmatrix}
\right) =
\begin{bmatrix}
  o_1 \\
  \vdots \\
  o_O
\end{bmatrix}
\] (2)

The formula (2) is the dot product of the \( O \times M \) dimension matrix and the \( M \) dimensional vector. The essence is linear transformation from one feature space to another, TCN Attention layer extract the features of sequence's emotion, so \( \vec{m} \) is in the space of emotional features, and the Dense layer maps the emotional feature space to the emotional category space. Finally, Dense layer output the probability of each sentiment category that the input belongs to.

4. TCN Attention Layer

A single model is hard to have good performance on specific tasks, because they are used to solve general problem. For specific scenario, these models may have acceptable performance, but have trouble to bring outstanding results. The common solution is to combine different models. This paper introduces the Attention mechanism to the TCN model and innovates a new model for Chinese sentiment analysis.

4.1. TCN

This paper chooses CNN to extract the emotional features of the sequence. CNN has excellent ability in feature extraction and is widely used in computer vision. However, CNN is not as good as RNN in natural language processing, for that CNN has never had a good model framework to process time series data. The emergence of TCN (Temporal Convolutional Networks) which making up for this vacancy [6].

TCN includes two structures: one-dimensional dilated convolution and residual module. One-dimensional dilated convolution operation is shown as formula (3). Each layer uses one-dimensional
dilated convolution kernel with extended rate of $d = 1, 2, 4$, which can cover all values of the input sequence.

$$y_s = \sum_{i=1}^{k-1} c_i \ast x_{s-d+i}$$  \hspace{1cm} (3)

The structure of the residual module is shown in Fig. 2, where the input is $\hat{z}^{l-1}$, let $F\left(\hat{z}^{l-1}\right)$ denote the operation of the two-layer convolution network, the output $\hat{z}^l$ is calculated as formula (4). Since the input form is different from the output, it is necessary to use $1 \ast 1$ convolution to modify the input dimension.

$$\hat{z}^l = F\left(\hat{z}^{l-1}\right) + \hat{z}^{l-1}$$  \hspace{1cm} (4)

The TCN structure has the following characteristics:

- One-dimensional dilated convolution used by the TCN network is a causal convolution. On the time series, causal convolution means that for the input on the $t^{th}$ time step, the convolution will only extract the feature of $j^{th}$ time step, where $j \leq i$. According to the structure of the TCN network, not only the causal relationship in the time dimension, but also the causal relationship between the convolutional layers. So, any output $y_t$ is related to the past input sequence $(x_0, x_1, \cdots, x_t)$, and not related with the future input sequence $(x_{t+1}, x_{t+2}, \cdots, x_T)$.

- The length of the model can be freely scaled to any length, which has the same structural advantages as the RNN model. Since the output of the TCN network is only related to the past inputs, its "input-output" structure can be designed in the form of $1 - n, n - 1, n - m$. This is very similar to the architecture of RNN.

In summary, the TCN can be considered to be a network structure consistent with RNN.

4.2. Add Attention Mechanism

In this section, we attempt to combine the TCN network with the Attention mechanism. Attention simulates the way that human brain works when processing information. When the human brain processes images, text or voice signals, we don't need to fully grasp any detailed information. Instead, when processing the information, the human brain will actively filter out the unimportant information, only leaving the most critical part.

The implementation of the Attention mechanism is divided into three steps:

**The first step** is to use a single layer perceptron to extract features from input $h_t$, as shown in the form (5), where $u_t$ is an implicit representation of $h_t$.

$$u_t = \sigma(W_t h_t + b_u)$$  \hspace{1cm} (5)

**The second step** uses a vector $u_w$ with randomized initialization multiply with $u_t$ to represent the importance of each input to the sequence, which obtains a normalized weight vector $a_t$ through Softmax regression. The calculation process is shown as formula (6).
\[ a_t = \frac{\exp(u_t^T u_{w0})}{\sum_{j=1}^{T} \exp(u_j^T u_{w0})} \]  

(6)

The third step is to use the weighted sum of the attention weight vector \( a_t \) and the input \( h_t \) as the output \( \vec{m} \), as shown in the formula (7).

\[ \vec{m} = \sum_{t=0}^{T} a_t u_t \]  

(7)

Finally, the TCN Attention structure for sentiment analysis model is shown in Fig. 3, the lower part of the network structure in the figure is the TCN model, and the upper part is the Attention mechanism. In the network structure, the TCN network acts as an Encoder to convert the input of each time step into corresponding features, and the Attention module is used to summarize the overall features.

Fig.3 TCN Attention structure for sentiment analysis

4.3. Model Optimization

4.3.1. Model Detail Optimization

The activation function used by the base TCN network layer is ReLU, which has the expression \( \max(0, x) \). Using ReLU as activation function can speed up the convergence of the model, but since the gradient in the negative interval of the function is 0, the dead ReLU problem [10] may occur, making the model very fragile. Therefore, this paper uses Leaky ReLU [11] function as the activation function of the model. The expression of this function is \( \max(\alpha x, x) \), where \( \alpha \) is a coefficient less than 1. Leaky ReLU function has all the advantages of ReLU, while avoiding the dead ReLU problem and enhancing the robustness of the model.

Secondly, convolutional layer of based TCN uses the same kernel size of the one-dimensional dilated convolution kernel. In this paper, the size of each convolution kernel is adaptively adjusted, so that the subsequent convolutional layer can obtain more abundant information.

4.3.2. Misidentification Problem Avoidance
Neural networks model is prone to over-fitting on data sets with obvious features, so it is relatively poor anti-interference ability for "noise". When analyzing neutral sentence, the result sometimes shows obvious emotional polarity. In this case, we called this phenomenon misidentification.

This paper solves this problem by expanding the training data set. Based on the existing data, we extend the data set vertically, and introduces the “neutral” emotional label in addition to the original “positive” and “negative” emotional labels. Based on the original data set, a total of 6,680 data that does not contain any emotional biases are expanded. These data generally include “news”, “daily language” and “comments without any feelings”.

After extending a classification, the emotion type label need to be encoded as One-Hot form. And the Output layer of the sentiment analysis model needs to use Softmax function to output the classification result, and the loss function needs to be changed from binary cross-entropy to categorical cross-entropy.

5. Evaluation

5.1. Data Set
Using a high-quality data set is a necessary condition for training excellent models. The main goal of this paper is to train a neural network model that can analyze the emotional polarity of Chinese E-commerce comments. Therefore, it is mainly necessary to collect positive product evaluation data (praise) and negative product evaluation data (bad reviews). The data needs to involve different kinds of goods, different sentence patterns and different moods as much as possible in order to ensure the accuracy and generalization of the model in different scenarios. Finally, this paper collects the dataset of: 10,679 positive comments and 10,428 negative comments. The comments covered: books, mobile phones, digital products, home appliances, and daily necessities. In this paper, 80% of the dataset is used as the training set for training model, and 20% of the dataset is used as the test set for evaluating the model.

5.2. Evaluation Indicators
Let the whole data set be U(x). For any sentiment category k, denote T_k(x) as the sequence set with label k, denote P_k(x) as the sequence set with label k predicted by the model, giving the definition of accuracy, precision, recall and F1-Measure, as is shown in formula (8) to (11), where accuracy shows the performance of the model on the whole dataset, while precision, recall, and F1-Measure are evaluation of the predicted results of the model on each category.

\[
\text{Accuracy} = \frac{\sum \lvert T_k(x) \cap P_k(x) \rvert}{\lvert U(x) \rvert} \tag{8}
\]

\[
\text{Precision} = \frac{\lvert T_k(x) \cap P_k(x) \rvert}{\lvert P_k(x) \rvert} \tag{9}
\]

\[
\text{Recall} = \frac{\lvert T_k(x) \cap P_k(x) \rvert}{\lvert T_k(x) \rvert} \tag{10}
\]

\[
F_1 - \text{Measure} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{11}
\]

5.3. Model Evaluation
The new model proposed in this paper is called TCN Attention V1, and the model after misidentification avoidance is called TCN Attention V2. LSTM, Bi-directional LSTM, LSTM Attention, TCN and the model we proposed are evaluated on the same dataset with following result.

The change of cross entropy and accuracy of each model are shown in Fig. 4.
Fig. 4 Cross entropy and accuracy trend for each model

The dashed line indicates the change of model's performance on the training set, and the solid line indicates the change of the model's performance on the test set. It can be found that after iterating several times on the training set, some model's cross entropy is no longer reduced, and the accuracy is no longer improved. After that, even worse performance will occur, at which point the model can be considered unsuitable for continued training. In the figure, Fig. 4a, 4b, and 4c are sentiment analysis models based on LSTM units. It can be found that after several iterations, the accuracy of these models on the test set is no longer improved, and the cross entropy drops to a certain value and starts to rebound. The accuracy rate on the training set is increasing, and the cross entropy is decreasing. This indicates that these models still have unstable factors on the data set used in this paper, and shows serious over-fitting phenomenon. However, all the proposed models (Fig. 4e and 4f) are stable after 6 iterations on the training set, and there is no cross entropy rebounding. The accuracy and cross-entropy of model does not appear much difference between training set and test set. In summary, the model proposed in this paper is more stable than other models.

The accuracy, precision, recall, and F1-Measure for each model on the test set are shown in the table 1.

The result of TCN Attention V2 model is better than other models in terms of accuracy, precision and F1-Measure of positive sentiment classification, precision, recall, and F1-Measure of negative classification. The comparison shows that the optimized model has been significantly improved.
Table 1. Evaluation for each model

| Model              | Acc  | Ppos | PNeg | Rpos | RNeg | F1pos | F1Neg |
|--------------------|------|------|------|------|------|-------|-------|
| LSTM               | 89.48| 89.77| 88.32| 89.90| 88.97| 89.83 | 88.64 |
| Bi-directional LSTM| 90.81| 88.43| 91.72| 92.53| 86.85| 90.43 | 89.21 |
| LSTM Attention     | 91.42| 91.50| 91.25| 91.75| 90.98| 91.62 | 91.11 |
| TCN                | 91.40| 92.81| 89.61| 91.16| 90.92| 90.38 |       |
| TCN Attention V1   | 92.62| 92.08| 92.74| 92.21| 91.60| 92.15 | 92.16 |
| TCN Attention V2   | **94.94**| **94.08**| **93.66**| **92.30**| **93.90**| **93.18**| **93.77**|

6. Conclusion
This paper combines TCN model with Attention mechanism to innovate a new model structure: TCN Attention model. It is proved that the over-fitting phenomenon effectively avoided by the model by horizontal comparison with LSTM, Bi-directional LSTM, LSTM Attention and TCN model. On test set data, TCN Attention V1 achieved accuracy: 92.62%, positive classification on F1-Measure: 92.15%, and negative classification on F1-Measure: 92.16%.

We further optimize TCN Attention V1 model, introducing third classification on the basis of the original emotion classification. The model has the ability to recognize noise, so that the extraction of emotion features is more accurate. On test set data, TCN Attention V2 finally achieved an accuracy rate of 94.94%, positive classification of F1-Measure: 93.18%, and negative classification of F1-Measure: 93.77%.

References
[1] Dodds, W.B., Monroe, K.B., Grewal, D.: Effects of price, brand, and store information on buyers’ product evaluations. Journal of marketing research (1991) 307–319
[2] Heaton, J.: Ian goodfellow, yoshua bengio, and aaron courville: Deep learning (2018)
[3] Hochreiter, S., Schmidhuber, J.: Long short-term memory. Neural computation 9(8) (1997) 1735–1780
[4] Liu, P., Qiu, X., Huang, X.: Recurrent neural network for text classification with multi-task learning. arXiv preprint arXiv:1605.05101 (2016)
[5] Mnih, V., Heess, N., Graves, A., et al.: Recurrent models of visual attention. In: Advances in neural information processing systems. (2014) 2204–2212
[6] Bai, S., Kolter, J.Z., Koltun, V.: An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. arXiv preprint arXiv:1803.01271 (2018)
[7] Hinton, G.E., et al.: Learning distributed representations of concepts. In: Proceedings of the eighth annual conference of the cognitive science society. Volume 1., Amherst, MA (1986) 12
[8] Mikolov, T., Sutskever, I., Chen, K., Corrado, G.S., Dean, J.: Distributed representations of words and phrases and their compositionality. In: Advances in neural information processing systems. (2013) 3111–3119
[9] Yang, Z., Yang, D., Dyer, C., He, X., Smola, A., Hovy, E.: Hierarchical attention networks for document classification. In: Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. (2016) 1480–1489
[10] Karpathy, A.: Cs231n convolutional neural networks for visual recognition. Neural networks 1 (2016)
[11] Xu, B., Wang, N., Chen, T., Li, M.: Empirical evaluation of rectified activations in convolutional network. arXiv preprint arXiv:1505.00853 (2015)