Chinese Microblog Sentiment Classification Based on Convolution Neural Network with Content Extension Method

Xiao SUN  
School of Computer and Information, Hefei University of Technology, Hefei, China 230009  
sunx@hfut.edu.cn

Fei GAO  
School of Computer and Information, Hefei University of Technology, Hefei, China 230009

Chengcheng LI  
IFLYTEK Co. Ltd.  
Hefei, China 230088

Fuji REN  
Tokushima University  
Tokushima, Japan 770800  
ren@is.tokushima-u.ac.jp

Abstract—Related research for sentiment analysis on Chinese microblog is aiming at analyzing the emotion of posters. This paper presents a content extension method that combines post with its’ comments into a microblog conversation for sentiment analysis. A new convolutional auto encoder which can extract contextual sentiment information from microblog conversation of the post is proposed. Furthermore, a DBN model, which is composed by several layers of RBM(Restricted Boltzmann Machine) stacked together, is implemented to extract some higher level feature for short text of a post. These RBM layers can encode observed short text to learn hidden structures or semantics information for better feature representation. A ClassRBM (Classification RBM) layer, which is stacked on top of RBM layers, is adapted to achieve the final sentiment classification. The experiment results demonstrate that, with proper structure and parameter, the performance of the proposed deep learning method on sentiment classification is better than state-of-the-art surface learning models such as SVM or NB, which also proves that DBN is suitable for short-length document classification with the proposed feature dimensionality extension method.

Keywords—DBN; RBM; ClassRBM; Social Network; Microblog Conversation; Sentiment Analysis.

I. INTRODUCTION

Micro-blog such as Sina Weibo or Twitter has been widely popularized in recent years. It acts not only as a way for interaction and communication among people, but also as a way to express or record individual emotions at work or in personal daily life. Microblog sentiment classification emerges as a challenging task. The emotion polarity of bloggers might reflect people’s hobbies and interests. Some related studies measure the public opinion such as preferences and political orientations of microblogger though analyzing the sentiment analysis of microblog text [1] [2].

The general procedure of emotion polarity detection commonly includes extraction of features from text and selection of machine learning models. The short length of microblog limits the feature dimension for sentiment analysis. Fortunately, posting microblog is a process of communicating with friends, so that comments for microblog are important reference information for analyzing the sentiment of a post in microblog. These comments of a post are adopted to expand feature dimensionality for sentiment analysis in microblog as the average length of Chinese microblog post is 13 Chinese characters on average subject to data and statistics. Short length of a post caused that traditional feature extraction method is hard to build reasonable representation for machine learning due to dimension sparse problem. The paper proposed a feature auto encoder, called Conversation to Sentence Convolutional Auto Encoder (ConCAE), to extract the context information of a post from microblog conversation. Furthermore, a DBN model is constructed by stacking a Classification RBM [3] [4] [5] [6] [7] [8] layer on top of several RBM layers together. Finally, we chose proper feature set and structure (include parameter) for the proposed model to perform some experiments and comparisons on Sina weibo corpus.

II. EXTENDED FEATURE EXTRACTION

The length of a post in Chinese microblog is limited to 140 characters, which causes the inefficiency of traditional feature extraction for sentiment analysis because of its brevity. In order to solve such problem, this paper proposes a post content extension method for feature dimension extension in Chinese microblog sentiment analysis problem. A post in microblog might be followed by several comments. These comments are public response to blogger, which reflect the emotion the post. The content of a post is the key factor in sentiment analysis and it’s comments are used as assistance features. We combine a post with it’s comments as a microblog conversation and extract sentiment-related feature by ConCAE. The content extension method employs some filtering approach to gain extended feature for microblog conversation composed by the post and it’s size-fixed comments. The first step captures the emotion and semantic information of words. It is supposed that every post and it’s comments is consisted by $m$ word $w_1, w_2, \ldots, w_m$, the word information $w_i$ is the integration of emotion information $w_{ei}$ and corresponding semantic information $w_{si}$. We composed size-fixed word bag and expression information to represent the feature of post and comments. Then post feature $V_{post}$ and comment feature sets $V_{com_1}, V_{com_2}, \ldots, V_{com_L}$ for a post can be obtained, $L$ is the number of comments. After word bag and expression information is obtained, auto encoding network is adopted to capture context information for the post and comments [9] [10] [11].
A. Word-level Feature Extraction

The aim of word feature extraction is to acquire emotional, semantic and integration information of words. The steps of word information extraction for a post are based on word segmentation and semantic analysis of the sentence. The emotional information of a word can be obtained from prior knowledge, such as a dictionary. Semantic information of each word in microblog is semantic role label, which can be obtained by analyzing the semantic structure of the microblog content.

The aim of word feature extraction is to acquire emotional, semantic and integration information of words [12]. The steps of word information extraction for a post are based on word segmentation and semantic analysis of the sentence. The emotional information of a word can be obtained from prior knowledge, such as a dictionary. Semantic information of each word in microblog is semantic role label, which can be obtained by analyzing the semantic structure of the microblog content.

1) Emotional Information of Words: Word is the smallest meaning expression units in Chinese, and Chinese expression composes of words and phrases. Therefore, the first step for Chinese sentiment analysis is word segmentation and part-of-speech(POS) tagging. All emotional words in blogs are marked by an external emotional dictionary in which each word is labeled with an emotional category and score \( w_{ei}(w) \), which represents the emotion information of the word.

2) Semantic Information of Words: The next step is semantic analysis of posts after Chinese word segmentation. Some modifiers of emotional words are also considered in the whole sentence of a post, which are listed in Table I. These modifiers refer to the interdependences between negative words, degree words and emotional words. Negative words mainly invert the semantic role of emotional words. We have collected 57 commonly used degree words and 37 negative seed words for the experiments. The seed sets are further expanded by some word-embedding methods such as Neural Network Language Model (NNLM, word2vec package by Google).

All degree words and negative words above are assigned with a prior weight in order to measure their impact. The value of degree words is called degree impact value (DIV), and the value of negative words is called negative impact value (NIV). Through constructing the semantic tree of a post, the degree words and negative words, which modify the polarity of emotional words in the post, can be labeled with its POS tag and semantic roles, the emotional value of an emotional word \( w \) will change. The semantic information of an emotional word can be calculated according to the following equation.

\[
 w_{si}(w) = \prod_{i=1}^{m} DIV_i \prod_{j=1}^{n} NIV_j
\]

Where \( m \) is the number of degree word and \( n \) is the number of negative words.

3) Integration Information of Words: Semantic information might transform the polarity of emotional word, which enhances or declines the emotional degree of an emotional word, so there is a meaningful relation between emotional information and semantic information for each word.

\[
 w_i = w_{ei} \cdot w_{si}
\]

Integration information takes emotion and semantic information of word into account for sentiment analysis of microblog, which specially consider the expression feature of a word under particular context.

B. Sentence-level Feature Extraction

For a post in microblog, it could receive varying amounts of comments from other microbloggers or friends. In order to extract enough feature from short text of a post and avoid the feature sparse problem. Some content extension methods should be considered. The content extension means extend length-limited post information by adopting size-fixed comment sets of the post in chronological order of the microblog conversation.

1) Microblog Conversation: For every post that need emotion recognition, the post and it’s streams of comments forms a microblog conversation. There is a filtering method which could validate whether a comment is a reply to the post or not. After filtering, the last \( L \) comments of the post is selected. We can get the post feature \( V_{post} \) and comments feature \( V_{com} \) according to section II. The comments feature sets \( V_{com}^{1}, V_{com}^{2}, \ldots, V_{com}^{L} \) could be transformed into comments matrix \( M_{coms} \), which could be seen on the left part of Figure 1. The column of \( M_{coms} \) is \( L \).
2) **Context Information of A Post:** Traditional method of extracting context information of a long sentence is unifying sentences information around the sentence with certain window size. The context information around a sentence usually has close semantic relations with each other. A microblog conversation(a post with its comments) belongs to group conversations. For the task of context information extraction of a post, the context information mainly reflects semantic links between post and comments, not interactions among comments which has weak semantic relations to each other. The paper settles this problem by proposing the ConCAE approach. The extracting process is described in Figure 1, the ConCAE approach produce feature information around post and its comments of a microblog conversation and then combines them using sampling method to get a context information representation(Matrix) of the post.

Given a microblog conversation which contains a post and L comments, we transform the microblog conversation to post feature \(V_{post}^L\) and comment feature sets \(V_{com}^1, V_{com}^2, ..., V_{com}^L\) which can be seen as a comments matrix. The input layer of the ConCAE is a post vector and streams of comment features. The convolutional layer combines inputs of \(V_{post}^L\) and \(V_{com}^1, V_{com}^2, ..., V_{com}^L\). The next step is to concatenating vector \(V_{con}\) for the post with its L comments:

\[
V_{con}^{l} = [V_{post}^{l}; V_{post}^{l}]
\]

\(L\) concatenated vector \(V_{con}\) can be obtained and stored in convolutional layer. The computing process of convolutional layer for each concatenation vector is as follows:

\[
v_{j}^{l} = f(W^{0}v_{con}^{l} + b^{0})
\]

Where \(W^{0}\) is the weight matrix for input vector \(V_{con}^{l}\) of convolutional layer. \(f(\cdot)\) is activation function for output such as sigmoid. \(b^{0}\) respects the bias terms. The convolutional layer shares the same weighting \(W^{0}\) and \(b^{0}\) with different input concatenation vector.

After the first convolutional layer, we can get a sequence of concatenation vector \(V_{1}^{l}, V_{2}^{l}, ..., V_{j}^{l}\). The sample layer would sample the same position of these concatenation vectors by byte, and then different weights for sampling points could be co-trained when ConCAE is stacked with other supervised learning model such as SVM.

### III. DBN MODEL INTRODUCTION

This paper builds a DBN model, which combines supervised learning with unsupervised learning, for Chinese microblog sentiment classification with extended features obtained by the method described in the former section. The DBN model is composed by a ClassRBM layer and several RBM layers [13] [14]. These RBM layers are used for achieve better representation of input feature. ClassRBM is used to calculate the conditional probability distribution of the label given input data. Besides directly achieving the label, ClassRBM also ensures that the features learning from the neural network are useful for classification.

#### A. RBM and ClassRBM

RBM is a typical two-layer neural network: a hidden layer \(H\) and a visual layer \(V\). The hidden layer and visual layer are fully connected with each other. The units in the same layer are independent to each other. The process of RBM training can be described in the following procedure: the feature vector of visual layer maps to the hidden layer, these visual layer units could be rebuilt by the hidden layer, these new visual units can remaps to a new hidden layer and then RBM would get new hidden units.

A commonly used RBM is Bernoulli-Bernoulli RBM which would satisfactorily handle binary and discrete feature space, but it is not applicable to continuous data. Meanwhile, the microblog feature introduced in the above sections is continuous. So Gaussian-Bernoulli RBM is adopted to solve the problem of continuous value input. Gaussian-Bernoulli RBM will change the binary variables of visual layer to follow Gaussian distributions, so that it can convert the continuous random variables to into binary random variables. Then Bernoulli-Bernoulli RBM can be used to handle binary input data. The first layer of multi-layer RBMs would adopt Gaussian-Bernoulli RBM and the rest layers are Bernoulli-Bernoulli RBM.

The energy function of Gaussian-Bernoulli RBM is defined as follows:

\[
E(v, h) = - \sum_{i=1}^{V} (v^i - a^i)^2/2\sigma^2 - \sum_{j=1}^{H} b_j h_j - \sum_{i=1}^{V} \sum_{j=1}^{H} w_{ij} v^i h^j \sigma_h^j
\]

The energy function of Bernoulli-Bernoulli RBM is:

\[
E(v, h) = - \sum_{i=1}^{V} a_i v_i - \sum_{j=1}^{H} b_j h_j - \sum_{i=1}^{V} \sum_{j=1}^{H} w_{ij} v_i h_j
\]

Where \(v\) is the visual input feature \((v_1, ..., v_n)\), with the hidden units \(h = (h_1, ..., h_m)\), \(\sigma\) is variances of input feature, and parameters \(\theta = (a, b, W)\) needs to be learned by training. For both kinds of RBM layers, it’s input feature is the combinations of input feature and hidden unit activations of previous layer. \(f(\cdot)\) is the activation function.

\[
h^{(k)} = f(u^{(k)} v^{(k)} + b^{(k)})
\]

\[
v^{(k+1)} = (u^{(k)}, h^{(k)})
\]

The next step is the self-training process of RBM based on CD (Contrastive Divergence) algorithm. First, the initial feature vector is assigned to the input layer, the next step is calculating the conditional probability distribution of the hidden layer to input layer; after that, the conditional probability distribution of the input layer to hidden layer is computed in the same way, then the input units are reassigned, the above steps are repeated according to preset threshold.

The difference between ClassRBM and RBM is that a sample class units \(Y\) is added. The input data of last network ClassRBM is composed by the sample feature representation vector and sample label vector \((0, 0, ..., 0)\). The first step is to redefine the energy function.

\[
E(y, v, h) = -c^T v - d^T h - h^T W v - e^T y - h^T U y - h^T U y
\]

Where the number of parameters becomes \(\Omega = (c, d, e, U, W)\). Mapping relations can be expressed as follows

\[
h = \text{sigmoid}(d + U y) + W v
\]
Where $sigm()$ is the same as sigmoid function $f()$. The conditional probability distribution among input layer and hidden layer or sample class could be obtained appropriately, but it hard to calculate their mutual joint probability distributions.

$$p(h|y,v) = sigm(d + U^T_y + Wv)$$  

$$p(y|h) = \frac{\exp(e_{y, v} + h^T U_{y,v})}{\sum_{y,v} \exp(e_{y, v} + h^T U_{y,v})}$$

The most commonly used training method is CD, which is also adopted in this paper. The stochastic gradient obtained by CD is shown in the following.

$$h_t = sigm(d + U_t \bar{Y}_t + W_t v_t)$$

$$h_{t+1} = p(h_t|y_t, v_t)$$

$$y_{t+1} = p(y_t|h_t)$$

$$v_{t+1} = p(v_t|h_t)$$

$$\Omega = \Omega - \lambda(E_{12}(y_t, v_t, h_t) - E'_{12}(y_{t+1}, v_{t+1}, h_{t+1}))$$

Where $\lambda$ is the learning rate. The DBN model is co-trained with ConCAE to get the optimal parameters.

### B. The Construction of DBN

The paper built a DBN model to complete emotional classification problem for microblog. The DBN model is composed by a single-layer ClassRBM and multi-layer RBMs. Figure 2 shows the structure of the DBN. First, multi-layer RBMs are trained layer by layer to optimize the input features like traditional RBMs training process to get prior weights for the final co-training. The output of previous RBM hidden layer is adopted as the input data for next RBM input layer. The feature representation of data in original space is transformed to another space by each layer of training, which improves the ability of feature representation for the final sentiment classification. The next step is the training of ClassRBM to get it’s prior weights. Then, the RBMs layers and Class-RBM layers are co-trained to tune the weights and get final parameters for classification. The output features of the top layer of RBM are set as the input data for ClassRBM layer. The classification layer is trained in the way of supervised learning and the discriminative performance of the feature representation is monitored in real-time. Initial concrete feature vector is converted into the feature vector in the abstract space. This transformation of feature expression optimizes the sample parameter feature and streamlines the whole training process.

DBN is the deep neural network with cascading Boltzmann model. The task of the each RBM layers is used to accomplish abstract representation of input features. The unsupervised pre-training process is constructed by bottom-up layer by layer. The later co-training process is to get the final optimal parameters for the whole DBN model.

### IV. EXPERIMENT AND RESULTS

In order to make comparisons, two different corpus are adopted in different language to validate the effect of content extension method: Twitter dataset and Sina microblog. The experiment prepares two sets data with different quantities. The Twitter datasets is adopted from International workshop on semantic evaluation 2013 task 2 corpora compiled by Vanzo et al. (2014), including 1391 labeled twitter with twitter comments [15] [16] [17] [18] [19]. The microblog database is crawled from Sina microblog and manually labeled with emotions(positive, negative or neutral). The collected microblogs with comments are adopted as analysis object and the number of comments of each microblog is over 6. The details of these two different datasets is shown in Table II. To verify the proposed methods, the experiment is designed for two purposes. The first is to prove that the method based on content extension feature extraction is more effective than single post feature extracting especially for short text. The second aim is to prove that DBN, which is composed by RBMs and ClassRBM, with better selected features and fine-tuned parameters, could further improve the performance of classification than traditional surface learning models [20] [21] [22] [23].

#### A. Measuring the Impact of Content Extension Method

Experiment 1 is aiming at measuring the impact of content extension method on microblog. For this purpose, there are different experimental sets with several different combinations of features mentioned above on the two databases. These optional features sets are shown in Table III. There are two

---

**TABLE II: Experimental Corpus**

<table>
<thead>
<tr>
<th>Set</th>
<th>Dataset</th>
<th>Class</th>
<th>Train</th>
<th>Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>D1</td>
<td>Twitter</td>
<td>Positive</td>
<td>455</td>
<td>106</td>
</tr>
<tr>
<td></td>
<td>(English)</td>
<td>Negative</td>
<td>200</td>
<td>52</td>
</tr>
<tr>
<td></td>
<td>Sina Microblog</td>
<td>Neural</td>
<td>460</td>
<td>118</td>
</tr>
<tr>
<td>D2</td>
<td>(Chinese)</td>
<td>Positive</td>
<td>1760</td>
<td>367</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Neural</td>
<td>3180</td>
<td>423</td>
</tr>
</tbody>
</table>

---


kinds of word-level features in Experiment 1: emotional information and integration information of words. For sentence-level feature, we adopted bag-of-word model to transform sentences into vectors, post-only information and post-context combination information by ConCAE are adopted as sentence-level features [24] [25] [26] [27]. Therefore four different basic feature sets are shown in Table III.

Table IV shows the result of different feature combinations with machine learning models on both data sets. There are four different feature combinations with two kinds of word information and two types of post feature. Meanwhile, SVM with optimal parameters (5-folder cross-violation) and one hidden layer DBN models (with one layer RBM taken 1500 hidden units) are adopted to perform the experiment. The size of system dictionary for bag-of-word is 2500.

The first general result is adopted as experiment baselines, in which the semantic information is not involved for word-level features and context information is not involved for post-level feature. The comparisons on different feature sets are also performed. The results show that the introduction of emotional word-level feature has no significant effects on the results, only considering the emergences of emotional words and ignoring semantic information of words. Almost all machine learning models are improved on the accuracy of classification which benefit from the addition of context information. It is also shown that KNN algorithmic is not sensitive to content extension. The classification effect of NB model performs better when the feature extracting method only considers statistics information. Instead of modest gain which had been expecting, recognition rates dropped 2% with the introduction of context information.

From the results shown in Table IV, only emotional feature cannot represent the semantic information well. The optimal feature set combination is Integration information with Post-Context information of post. SVM and DBN model are more effective in sentiment classification task.

In the next experiment, the purpose is to select the proper dimension of post and comment feature which is the size of system dictionary that bag-of-word used. Four experiments are set to get the optimal dimension value:

a) This experiment chooses second feature set(W2+P1) as the initial feature and the feature is adopted in training SVM;
b) This experiment combines Integration information feature and Post-Context information feature as the initial feature which is adopted in training SVM;
c) This experiment chooses second feature set(W2+P1) as the initial feature for training, the model is DBN, which includes a RBM layer and a ClassRBM layer;
d) This experiment combines Integration information feature and Post-Context information feature as the initial feature. The DBN is composed by a RBM and a ClassRBM layer.

In these four experiments, for the sake of comparisons, we set the optimal parameters for two models by 5-folder cross-violation. The data set D2 is the experiment dataset. The optimal parameter of SVM is \((\gamma = 3, \text{coef} f = 1.5)\). The number of hidden units of DBN is one layer RBM with 1500 hidden units.

Figure 3 shows the experimental results, as can be seen from these four sub-figures, general trends are mostly the same: the accuracy increases with the feature dimension value(from 1000 to 3000), accuracy grows fast when the dimension value is less than 2500 and slowly after 2600. When the value is under 1500, all the systems perform badly; the feature vector cannot represent the short microblog text well because of inadequate dimension. The reason why only one RBM layer is picked here is that the calculation complexity increases with the feature dimension value. Taking accuracy and computation complexity into account, the final experiment chose 2600 as the text-based modality feature dimension. Comparing Fig.3(a), Fig.3(b), Fig.3(c) and Fig.3(d), the performance of feature extraction method is better than traditional feature extraction method. When the dimension value is limited to 1000, feature extraction got more features than traditional feature extraction method to obtain better accuracy, which is reasonable for application. For further proving the effectiveness of the proposed content extension method, proper context information of post from microblog conversation with different size of comments should be obtained. Figure 4 show the results with different number of comments for the context information extraction of post. These results indicate a correlation between context information of post and the size of comments. As comment size increases, the accuracy of classification increases until the size reaches 7. The accuracy decreases with the size increase after the number over 9. The comments are filtered by time. According to the increment scale of the conversation, more noise data sets might be introduced which might directly bring on the decrease of recognition results.

B. Optimal Shapes of DBN Model

For the second purpose, the proper shapes and the depth of DBN should be determined. The results for various DBN shapes (seen in Figure 5) are referred by the number of hidden units in the top level of DBN. The DBN structure adopted in this paper is one top classRBM layer. It can be seen here is
that the DBN shape has little impact on the performance of DBN. When the number of hidden units (of class RBM) exceeds 500, the accuracy remains almost the same around 76% and then DBN shape slightly impact the performance of DBN in the following curve. Several experiments of DBN model with different depth are designed to further verify the effectiveness of deep network model for classification and to choose the optimal depth of DBN. We set the maximum depth as 7. These numbers of hidden units from bottom up are 2000, 1500, 1000, 1000, 500, 500, 500 respectively.

It can be seen from Figure 6 that precision maintains a steady growth trend with growth of DBN depth, although the recall of them has undulatory property. The precision gradually declines while the depth of RBM layers surpasses 4 layers. When the layers of RBM reaches 7, precision tends to be lower than DBN model with less layers, which might be caused by insufficient training data. The complexity of nonlinear functions that need to be constructed grows with the increasing complexity of DBN model. Accordingly, the energy losses in such functions increase. The feature information might be lost during optimization, which might in return cause the decreased accuracy.

V. CONCLUSIONS AND FUTURE WORK

In this paper, we proposed a deep belief nets (DBN) model with content extension method ConCAE to solve features sparse problem caused by the limitation of the length of short text in Chinese microblog for sentiment classification. The results demonstrate that the proposed post context information extended and feature extraction method is more effective than traditional feature extraction method, and the performance of deep learning with proper structure and fine-tuned parameter on sentiment classification is better than traditional surface learning models such as SVM or NB. The performance could be further improved in the following aspects: all the relationships between bloggers could be calculated, which might also increase the computational complexity; Meanwhile, the system time-consuming increases with the number of DBN layers during training, which could be optimized by adapted some faster training and feature cutting algorithms.

ACKNOWLEDGMENT

The work is supported by National Natural Science Funds for Distinguished Young Scholar (No.61203315) and the Natural Science Foundation of Anhui Province (No.1508085QF119), and also supported by the State Key
Program of National Natural Science of China (No.61432004). This work was partially supported by JSPS KAKENHI Grant Number 15H01712.

REFERENCES


