Microblog Sentiment Analysis with Emoticon Space Model

Fei Jiang[1], Yiqun Liu[1,*], (Senior Member, CCF), Huanbo Luan[1], Jiashen Sun[2], Xuan Zhu[2], Min Zhang[1] (Member, CCF), Shaoping Ma[1]

[1]*State Key Laboratory of Intelligent Technology and Systems, Tsinghua National Laboratory for Information Science and Technology, Department of Computer Science and Technology, Tsinghua University, Beijing 100084, China*

[2]*Language Computing Lab, Samsung R&D Institute of China, Beijing 100028, China*

E-mail: f91.jiang@gmail.com, yiqunliu@tsinghua.edu.cn, luanhuanbo@gmail.com, jiashens.sun@samsung.com, xuan.zhu@samsung.com, z-m@tsinghua.edu.cn, msp@tsinghua.edu.cn

Received November 15, 2014

Revised May 16, 2015

Abstract Emoticons have been widely employed to express different types of moods, emotions and feelings in microblog environments. They are therefore regarded as one of the most important signals for microblog sentiment analysis. Most existing works use several emoticons that convey clear emotional meanings as noisy sentiment labels or similar sentiment indicators. However, in practical microblog environments, tens or even hundreds of emoticons are frequently adopted and all emoticons have their own unique emotional signals. Besides, a considerable number of emoticons do not have clear emotional meanings. An improved sentiment analysis model should not overlook these phenomena. Instead of manually assigning sentiment labels to several emoticons that convey relatively clear meanings, we propose the emoticon space model (ESM) that leverages more emoticons to construct word representations from a massive amount of unlabeled data. By projecting words and microblog

*Corresponding author

This work was supported by Tsinghua-Samsung Joint Lab, National Key Basic Research Program (2015CB358700) and the National Natural Science Foundation of China under Grant No. 61472206, 61073071, and 61303075.

posts into an emoticon space, the proposed model helps identify subjectivity, polarity and emotion in microblog environments. The experimental results for a public microblog benchmark corpus (NLP&CC 2013) indicate that the ESM effectively leverages emoticon signals and outperforms previous state-of-the-art strategies and benchmark best runs.

Keywords Microblog Sentiment Analysis, Emoticon Space, Polarity Classification, Subjectivity Classification, Emotion Classification

1 Introduction

Microblogs, such as Twitter and Sina Weibo, are a popular social media in which millions of people express their feelings, emotions, and attitudes. Because a large number of microblog posts are generated every day, the mining of sentiments from this data source assists in the performance of research on various topics, such as analyzing the reputation of brands [1], predicting the stock market [2] and detecting abnormal events [3]. Therefore, an improvement in the performance of sentiment analysis tasks in microblog environments is crucial. Microblog sentiment analysis has been a hot research area in recent years, and several important issues have been studied, such as identifying whether a post is subjective or objective (subjectivity classification) [4, 5], identifying whether a post is positive or negative (polarity classification) [4, 5], and recognizing the emotion in a particular post (emotion classification) [3].

Supervised machine learning techniques have been widely adopted to microblog sentiment analysis and have proven to be effective [6]. Various features, such as the sentiment lexicons, the part-of-speech tags, and microblogging features have been exploited to reinforce the classifiers [7, 8]. However, manually labeling sufficient training posts is extremely labor intensive because of the large vocabulary adopted by microblog users. Fortunately, in microblog environments, various emoticons are frequently adopted and are usually posted along with emotional words. In our investigation, there are more than a thousand different emoticons in Sina Weibo, and microblog posts containing emoticons take up a proportion of over 25% in a dataset of about 20 million posts randomly collected from March 2012 to December 2012 (Dataset 1). Moreover, graphical emoticons, which are more vivid compared with those composed of punctuation marks, have been introduced in many microblog platforms. Thus, emoticons can serve as an effective source of emotional signals, making it possible to perform sentiment classification tasks without or with a small amount of manually labeled posts.

A small portion of emoticons has very clear emotional meanings, such as  and  for positive sentiment,  and  for negative sentiment. Many existing works use these emoticons as noisy sentiment labels of microblog posts to train classifiers [9, 10, 3] and to avoid the expensive cost of manually labeling data. However, the performance of these research efforts is affected by several problems. First, as contexts and user habits vary, the sentiment of the words in such posts may not be the same as the emoticons. Thus, noises are introduced. Second, every emoticon has a unique emotional meaning in microblog environments. For example,  and  both express an emotion of happiness, but words that co-occur with the latter tend to be more naughty. Such information can be not utilized by equally treating them as the noisy positive label. Moreover, some emoticons, such as  and  do not have clear emotional meanings and it is hard to assign sentiment labels to the emoticons. They form a considerable portion of the employed emoticons and posts that contain these emoticons have a proportion of over 7% in Dataset 1. These emoticons may also help identify sentiment. In our investigation, the emotion of emoticon  (Ultraman) is uncertain, but posts containing this emoticon are unlikely to have an emotion of anger.

For the first problem, a state-of-the-art method has been proposed to combine noisy data and manually labeled data to effectively train classifiers [5]. However, no previous models have been proposed to fully exploit the potential of emoticons, by either differently treating each emoticon or integrating those without clear emotional meanings.

To better utilize the signals contained in emoticons and to improve the performance of microblog sentiment classification tasks, we propose a semi-supervised model, called the emoticon space model (ESM). In this model, we select a relatively large number of commonly employed emoticons with and without clear emotional meanings to construct an emoticon space, where each emoticon serves as one dimension. Typically, the ESM consists of two phases: the projection phase and the classification phase. In the first phase, posts are projected into the emoticon space based on the semantic similarity between words and emoticons, which can be learned from a massive amount of unlabeled data. In the second phase, the assumption is made that posts with similar sentiments have similar coordinates in this space. Therefore, supervised sentiment classifiers are trained using the coordinates of the posts as features. By this means, sentiment analysis tasks, such as subjectivity classification, polarity classification and emotion classification can be well performed. Because the dimension of the emoticon space is considerably lower than that of the word space (i.e., the size of the vocabulary), the supervised classification tasks in emoticon space can be performed using less manually labeled data.

2 Related Work

There are several well defined tasks in microblog sentiment analysis. The most commonly studied tasks include document level classification tasks (subjectivity classification, polarity classification, emotion classification). Moreover, Weichselbraun et al. [11] concentrated on concept level sentiment analysis. They tried to infer the sentiment of the concepts with data and existing sources such as WordNet. Jiang et al. [4] proposed to recognize the sentiment towards a given object. In this paper, we focus on the document level classification tasks.

Both supervised and unsupervised methods have been used for microblog sentiment analysis. The construction of unsupervised classifiers based on existing sentiment lexicons is an option as labeled data are not required [6, 12]. However, Bermingham et al. [6] determined that a classifier based on SentiWordNet [13] performed poorly compared with supervised methods. Other unsupervised methods based on emoticons or sentiment lexicons were proposed by Hu et al. [14] and Cui et al. [15]. Kouloumpis et al. [7], and Barbosa et al. [8] used sentiment lexicons as features to reinforce supervised classifiers.

Supervised methods are effective for microblog sentiment analysis [6, 4]. Birmingham et al. [6] utilized the multinomial naive Bayes (MNB) and support vector machine (SVM) to perform supervised sentiment classification. First, several thousand training posts were manually labeled by a team of nine researchers. The limitation of supervised methods is that the performance is highly dependent on the size of the manually labeled data, which are always labor intensive to obtain.

Emoticons are frequently used to alleviate the problem. Pak et al. [9] used emoticons that have relatively clear emotional meanings as noisy labels of posts. They used ":)", and ":-)" as the noisy positive labels and ":-(" and ":(" as the noisy negative labels. Posts that contain these emoticons were used as training data, and an MNB classifier was trained for polarity classification. Similar work has been performed by Bifet et al. [16] and Go et al. [10]. Zhao et al. [3] used emoticons as noisy labels to perform emotion classification on Sina Weibo. Hu et al. [14] verified sentiment indication of emoticons. Other signals, such as hashtags, have also been used as noisy labels [7, 17]. We refer to these types of methods as noisy supervised methods for the remainder of this paper.

Additionally, noisy supervised methods are adopted to subjectivity classification. Pak et al. [9] regarded the emoticons listed above (":)", ":(", etc.) as the noisy subjective labels. To acquire noisy objective data, they assumed that posts from accounts of popular newspapers and magazines are objective. In addition, Liu et al. [5] assumed that posts containing an objective url link are objective: If a url link does not represent pictures or videos, then this link is an objective url link.

Noisy supervised methods are negatively affected by the noises of training data. Liu et al. [5] proposed a novel method called the emoticon smoothed language model (ESLAM) to overcome the shortcomings of both supervised methods and noisy supervised methods.

As stated before, all previous works can not make full use of the rich emoticon signals, which motivates us to take a further step in this direction.

3 Emoticon Space Model

In ESM, we use a number of frequently adopted emoticons to construct an emoticon space. As mentioned in the *Introduction* section, ESM consists of two phases: the projection phase and the classification phase. For the projection phase, to satisfy our assumption that posts with similar sentiments have similar coordinates in the emoticon space, we obtain the coordinates of words using the semantic similarity between words and emoticons. Afterwards, the coordinates of the posts are obtained based on coordinates of words. For the classification phase, we use the coordinates of the posts as features for supervised sentiment classification tasks.

In this section, we first introduce a distributed representation of words [18], which provides an effective way to learn the semantic similarity between words and emoticons. Based on the distributed representation, we then present how words and posts are projected into an emoticon space. Finally, we introduce how the supervised sentiment classification is performed using the coordinates of the posts.

3.1 Distributed Representation of Words

The distributed representation of words has been widely used in neural probabilistic language models (NNLMs). In a distributed representation, each word is represented by a continuous real-value vector [18]. In this paper, words that are used in similar contexts are considered semantically similar and tend to have similar vectors [19]. The vectors can be learned by using a massive amount of unlabeled data and will be used later to project words into the emoticon space.

We use *word2vec* [20, 21] to learn the distributed representation of words, because of its fast training speed.

3.2 Word Projection

When learning a distributed representation, we preprocess the corpus and treat each emoticon as a special text token. For example, we use *tkn-emoji-laugh* to denote the emoticon 😁. If not specified, words in the rest part of this paper include emoticons.

By leveraging large amounts of microblog corpora, the representation vectors of all words are learned by *word2vec*. The representation vectors of words form a matrix $\boldsymbol{M_w} \in \mathbb{R}^{d \times V}$, where V is the size of the vocabulary and d is the dimension of the representation vectors. Each column of $\boldsymbol{M_w}$ denotes the representation vector of the corresponding word.

Suppose E emoticons (denoted as $(e_1, e_2, ..., e_E)$) are selected to construct the emoticon space. We search for the representation vectors of these emoticons in matrix $\boldsymbol{M_w}$ and receive a matrix $\boldsymbol{M_e} \in \mathbb{R}^{d \times E}$. $\boldsymbol{M_e}$ is a sub matrix of $\boldsymbol{M_w}$. Each column in $\boldsymbol{M_e}$ denotes the representation vector of the corresponding emoticon.

In the distributed representation, words used in similar contexts tend to have similar vectors. Therefore, measuring the similarity between the representation vectors of word w_i and emoticon e_j helps identify their semantic similarity. In this paper, the cosine distance is used as the measurement of similarity between the representation vectors, which can be formalized as equation 1.

$$similarity(\vec{\boldsymbol{w_i}}, \vec{\boldsymbol{e_j}}) = \frac{\vec{\boldsymbol{w_i}} \cdot \vec{\boldsymbol{e_j}}}{|\vec{\boldsymbol{w_i}}| \, |\vec{\boldsymbol{e_j}}|} \qquad (1)$$

where $\vec{\boldsymbol{w_i}}$ and $\vec{\boldsymbol{e_j}}$ are the representation vectors of w_i and e_j. Specifically, if $w_i = e_j$, the similarity between the representation vectors is 1. We simply use equation 1 as the measurement of semantic similarity between w_i and e_j and use the semantic similarity as the coordinate of the word w_i in dimension j. Algorithm 1 shows the process of calculating the coordinate matrix of all words, which is denoted by $\boldsymbol{C}$.

In Algorithm 1, $\boldsymbol{C} \in \mathbb{R}^{E \times V}$, and each column of $\boldsymbol{C}$ represents the coordinate of the corresponding word or emoticon in the emoticon space. Thus, words have been projected into the emoticon space. If a word refers to an emoticon which is used to construct the emoticon space, then the corresponding column of $\boldsymbol{C}$ can be considered as the basis of this emoticon space. Because different emoticons are interrelated, this emoticon space is a non-orthogonal space.

3.3 Microblog Post Projection

In the last section, we have proposed a simple method to project words into the emoticon space using the semantic similarity between words and emoticons. The semantic similarity between

Algorithm 1 Calculation the coordinates of words and emoticons

Require:

Distributed representation matrix of words and emoticons, $\boldsymbol{M_w}$

Distributed representation matrix of emoticons, $\boldsymbol{M_e}$

for each i in $[1:E]$ **do**
  for each j in $[1:V]$ **do**
    $\boldsymbol{C}(i,j) = similarity(\boldsymbol{M_w}(:,j), \boldsymbol{M_e}(:,i))$
  end for
end for

posts and emoticons, however, can not be learned directly, but the coordinates of the posts can be obtained using basic mathematical operations on the coordinates of the words.. In this paper, we investigate two simple strategies for post projection.

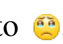
- **Basic ESM (B-ESM)**

  The simplest way to project a particular post into the emoticon space is to sum up the coordinates of the words that form the post. Formally, let p be the post, and $\vec{p}$ be the coordinate of the post. Therefore, $\vec{p}$ can be computed with equation 2.

$$\vec{p} = \sum_{w_j \in p} \boldsymbol{C}(:,j) \tag{2}$$

  We name this strategy as Basic ESM, for its simpleness. Benefiting from the property that the coordinates of words are bounded in $[-1,1]$, each word only has a limited effect for post projection.

- **Extended ESM (E-ESM)**

  From our observation, many subjective posts contain one sentiment word or several sentiment words. Sentiment words may be semantically similar to some emoticons, and semantically dissimilar to other emoticons. For example, the word *happy* may be semantically similar to , and semantically dissimilar to . Thus, the coordinates of the word *happy* for the corresponding dimensions may be relatively large or small. Therefore, the maximum and minimum values of word coordinates in certain dimensions of a particular post may indicate the sentiment of this post.

  ESM is flexible to integrate this information. Based on B-ESM, we add the maximum and minimum values above to the posts. Therefore, the coordinate of a particular post in each dimension can be

$$\vec{p} = \begin{bmatrix} (\min\limits_{w_j \in p} \boldsymbol{C}(1,j), \sum\limits_{w_j \in p} \boldsymbol{C}(1,j), \max\limits_{w_j \in p} \boldsymbol{C}(1,j)) \\ (\min\limits_{w_j \in p} \boldsymbol{C}(2,j), \sum\limits_{w_j \in p} \boldsymbol{C}(2,j), \max\limits_{w_j \in p} \boldsymbol{C}(2,j)) \\ \cdots \\ (\min\limits_{w_j \in p} \boldsymbol{C}(E,j), \sum\limits_{w_j \in p} \boldsymbol{C}(E,j), \max\limits_{w_j \in p} \boldsymbol{C}(E,j)) \end{bmatrix} \tag{3}$$

Table 1. Statistics of the NLP&CC 2013 Benchmark Dataset

	neutral	like	happiness	sadness	disgust	anger	surprise	fear
training set	1823	597	371	388	425	235	112	49
testing set	4925	1525	1106	759	969	405	221	90
total	6748	2122	1477	1147	1394	640	333	139

denoted as a triplet of real values. The coordinate of the post can be formalized as equation 3.

3.4 Supervised Sentiment Classification

After projecting posts into the emoticon space, the supervised sentiment classification tasks can be performed by using the coordinates of the posts as features. For B-ESM, the coordinates are used as feature vectors directly. For E-ESM, triples in all dimensions are concatenated to form a feature vector. The advantage of ESM is that emoticon signals are fully leveraged, and that it performs the supervised sentiment classification in an emoticon space in which the dimension is much lower than the word space. The next section will illustrate the performance of the two ESM models for the sentiment classification tasks.

4 Experiments

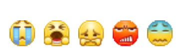
4.1 Experiment Setups

Experimental studies are performed on a public Chinese microblog benchmark corpus (NLP&CC 2013). This corpus consists of both a training set and a testing set[1,2,3]. The dataset

1. http://tcci.ccf.org.cn/conference/2013/dldoc/evsam02.zip
2. http://tcci.ccf.org.cn/conference/2013/dldoc/evdata02.zip
3. http://tcci.ccf.org.cn/conference/2013/dldoc/evans02.zip

is composed of fourteen thousand microblog posts collected from Sina Weibo and each of them is annotated with one of the following eight emotion tags: neutral, like, happiness, sadness, disgust, anger, surprise and fear. Posts in the dataset are not limited to certain topics. Table 1 shows details of the benchmark corpus. This dataset is adopted for the evaluation of polarity classification, subjectivity classification and emotion classification tasks. In polarity classification, emotions such as happiness and like are regarded as positive sentiment while sadness, anger, disgust and fear are regarded as negative sentiment[1]. In subjectivity classification, neutral is regarded as objective and the other seven emotion types are regarded as subjective. In emotion classification task, each emotion type serves as one class.

Table 2. Emoticons with clear emotional meanings

Sentiment	Example emoticons	Amount
positive		33
negative		19
happiness		9
like		10
sadness		7
disgust		5

To evaluate the performance of the proposed ESM model, 100 emoticons[2] which are the most frequently used in microblog posts are selected to the construct the emoticon space. ICTCLAS [22] is adapted to the microblog corpus for Chinese word segmentation. LIBSVM [23] is chosen for supervised classification in ESM, and for baseline methods when SVM is needed. Dataset 1 is used for the projection phase of the ESMs.

Three state-of-the-art baseline strategies are adopted and compared with the proposed ESM framework for polarity, subjectivity and emotion classification tasks. Sufficient data with noisy labels are provided for the baseline methods and the parameters of the baseline methods are well tuned to achieve their best performance.

- **Supervised methods**: These methods use manually labeled data to train classifiers. We investigate MNB and SVM, which are the most widely adopted classifiers for these tasks. We use binary feature vectors for SVM similar to [6]. We abbreviate supervised MNB and SVM as S-MNB and S-SVM, respectively.

- **Noisy supervised methods with emoticons**: We use emoticons that have clear emotional meanings from the 100

1. We do not use emotion of surprise for polarity classification, as the polarity of this emotion is undetermined in this dataset.

2. Except for the results in Figure 2.

emoticons mentioned above as noisy sentiment labels. Table 2 shows the five most frequently adopted emoticons for each sentiment. The assumption of [9] is adopted for the noisy objective label and outperforms the assumption of [5]. We implement a noisy MNB (i.e., N-MNB) classifier, which is adopted by most previous works.

- **Combination methods**: [5] propose a state-of-the-art method that effectively integrates the supervised methods and the noisy supervised methods. Their method is called the ESLAM model. ESLAM is primarily used to identify polarity and subjectivity, but can be easily extended to perform emotion classification by taking emoticons that clearly express the corresponding emotions as noisy labels.

To better explore the benefits of differential treatment for each emoticon and the effect of emoticons that do not have clear emotional meanings, the E-ESM(*) and the B-ESM(*), which use the same emoticons as the baseline methods, are investigated. The accuracy of all methods is reported to measure the performance of polarity, subjectivity and emotion classification tasks.

We then study a widely adopted special case of polarity classification: the lexicon-based polarity classification, which does not require manually labeled posts. Finally, we illustrate the improvement of the ESMs over the best runs in the NLP&CC 2013 benchmark.

Finally, we will explore whether the emoticons are irreplaceable when constructing such a sentiment space. More specifically, we will replace the emoticons with sentiment words to construct the sentiment space. We call this model Word Space Model (WSM) which is similar to Emoticon Space Model. We will compare the performance of WSM and ESM. As resources are limited, this experiment is only done for polarity classification.

4.2 Polarity Classification

For polarity classification, the original dataset has 3,599 positive posts and 3,320 negative posts. Next, a balanced dataset of 6,640 posts is randomly sampled, using the 3,320 negative posts and a random sample of 3,320 posts from the 3,599 positive posts. Similar to [5], a balanced subset of X total posts is randomly sampled for training along with 6,640 - X posts for testing. This procedure is performed for a given X and fixed parameters in ten rounds. The subjectivity classification and the emotion classification below use the same evaluation methodology. Classification accuracies for a different size of the manually labeled training set

X are explored. In this task, X varies as 250, 500, 1000, 2000, 3000, 4000 and 5000.

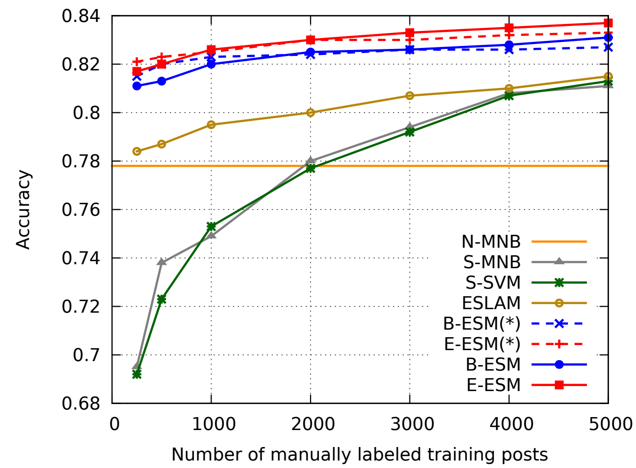


Fig. 1. Accuracies for different numbers of manually labeled training posts in polarity classification.

In this task, the SVM with a linear kernel achieves the best performance for the ESMs. The polarity classification results of the ESMs and the baseline methods are compared in Figure 1. From this figure we can see that the B-ESM(*) and the E-ESM(*) outperform all baseline methods, and indicate that the use of the same emoticons for ESM results in an improved utilization of emoticon signals. Emoticons that do not have clear meanings are useful when the number of training posts becomes relatively larger but have a negative impact when the training data are insufficient. The ESMs consistently outperform baseline methods for different training dataset sizes. According to the statistics, the E-ESM slightly outperforms the B-ESM because more features of the posts are introduced. When the training dataset size is larger than 4,000, the performance differences between the ESLAM and the supervised methods (S-MNB and S-SVM) become closer. However, a large gap still exists between the baseline methods and the ESMs.

Note that the ESMs are very robust even when the training size is small. The E-ESM with a training size of 250 outperforms all baseline methods with a training size of 5,000. For only a small number of manually labeled posts, a relatively high accuracy for the ESMs is obtained.

We then investigate the performance of the B-ESM and the E-ESM using different numbers of emoticons. For a given number N, N emoticons that are most frequently adopted in Sina Weibo are selected for the ESMs to construct the emoticon space. Among the N emoticons, ones that have clear polarity indications are used as noisy sentiment labels for the ESLAM. The models are trained with 5,000 manually labeled posts. Figure 2 shows the performance of the models as N varies. We can see that with more emoticons adopted, the ESMs achieve a better performance and outperform the ESLAM. Moreover, the ESMs are flexible to leverage much more emoticons at no extra cost except for a few more computations.

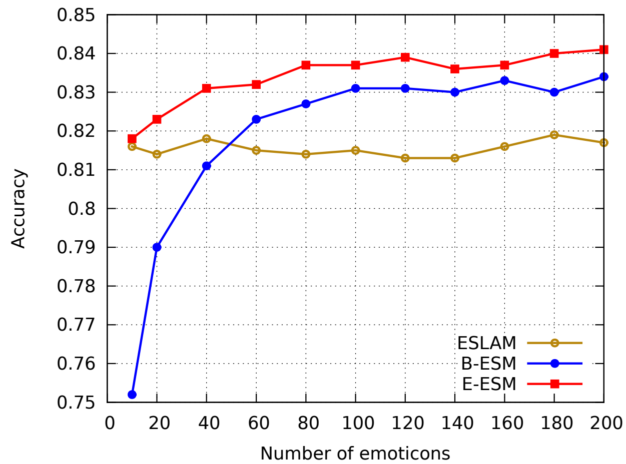


Fig. 2. Accuracies of the ESMs and the ESLAM using different numbers of emoticons in polarity classification.

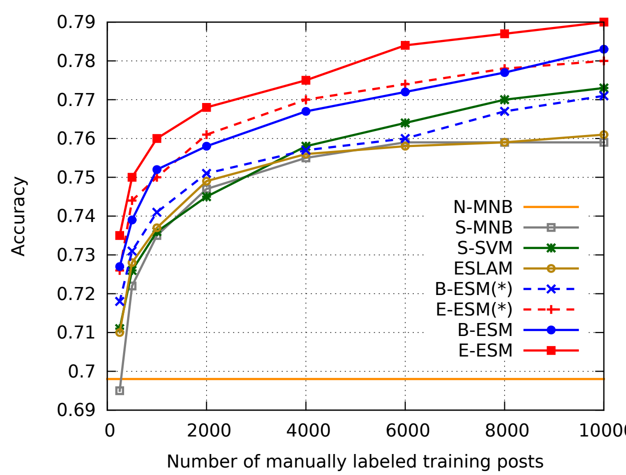


Fig. 3. Accuracies for different numbers of manually labeled training posts in subjectivity classification.

4.3 Subjectivity Classification

Similar to polarity classification, we randomly sample a balanced dataset of 13,496 posts for subjectivity classification. In this task, the size of the manually labeled training set varies as 250, 500, 1000, 2000, 4000, 6000, 8000 and 10000. The results are shown in Figure 3.

For baseline methods, when the training size is larger than 4,000, the S-SVM outperforms both the S-MNB and the ESLAM, and even slightly outperforms the B-ESM(*). However, the other three ESMs outperform all baseline methods consistently, for different training sizes. All ESMs outperform the noisy supervised methods and the ESLAM. Meanwhile, after comparing the E-ESM with the E-ESM(*) and the B-ESM with the B-ESM(*). We can see that emoticons that do not have clear emotional meanings help subjectivity classification.

In subjectivity classification, the ESMs use more training data to obtain a relatively high performance when compared with polarity classification. However, to achieve a comparable accuracy, the B-ESM and the E-ESM still require less labeled data than the baseline methods.

4.4 Emotion Classification

We use four emotion types (like, happiness, sadness and disgust) for a maximum number of posts to perform this task because the quantities for the other three types are relatively small. Similar to polarity classification, a balanced set of 4,588 posts is sampled and contains 1,147 posts for each type. The size of the manually labeled training set varies as 400, 800, 1600 and 3200.

Figure 4 shows the results for this task. The results are similar to polarity classification. the E-ESM outperforms the E-ESM(*) when the training size is no less than 800 similar to the B-ESM. The ESMs outperform all baseline methods for different training sizes and are robust and achieve high performances when the training size is relatively small. For example, the E-ESM(*) achieves an accuracy of 0.626 for 400 manually labeled training posts, which is higher than the best performance (0.617) of the baseline methods with a training size of 3,200.

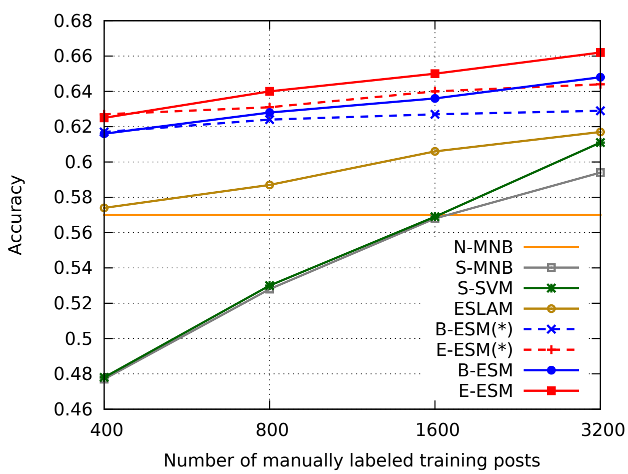


Fig. 4. Accuracies for different numbers of manually labeled training posts in emotion classification.

4.5 Lexicon-Based Polarity Classification

Because words can be considered as special posts, ESM can use raw sentiment words as training data to identify the polarity of the posts. We randomly sample 125 positive words and 125 negative words which are contained in Dataset 1 after word segmentation from [24] as the training data, and use the 6,640 posts above as the testing data for the B-ESM(*). This procedure is performed for fixed parameters in ten rounds. The B-ESM(*) achieves an accuracy of 0.803, which slightly outperforms the ESLAM with a training size of 2,000. As a comparison, only 6% of the testing posts contain at least one of these words and a traditional lexicon matching method [6] based on the 250 words only achieves an accuracy of 0.518. We conclude that ESM not only inherits the advantage of the lexicon-based methods that manually labeled data are not required, but also achieves a high performance.

4.6 Comparison to Word Space Model

The main idea of ESM is to learn a sentiment representation of words, with the help of emoticons. Intuitively, we can use sentiment words instead of emoticons to do similar things.

The sentiment lexicon, [24], contains 782 positive words and 942 negative words in total. After filtering out the words which are not contained in our unsupervised corpus (Dataset 1), we finally get 699 positive words and 883 negative words.

We use Extended Word Space Model (E-WSM, which is just similar to E-ESM) and randomly select 100, 200, 500 and all words respectively to construct the sentiment word space. We then compare the performance of E-WSM and E-ESM in polarity classification.

Figure 6 shows the results for this task. We can see that the E-WSM models are only comparable to our baseline model ESLAM. Moreover, 100 words are enough for the E-WSM models and using more sentiment words does not have significant improvement.

Given the big performance gap between E-ESM and E-WSM, we conclude that for microblog sentiment analysis, or at least polarity classification, emoticons are irreplaceable to learn similar representations. We believe that this is because emoticons are purer carriers of sentiment, while words have much more complex senses and contexts, such as negations, ironies, etc.

4.7 Visualization

Following the previous subsection, we acquire the coordinates of the sentiment words in emoticon space and then visualize the emoticon space by performing dimensionality reduction on the coordinates using Principal Component Analysis (PCA). The number of components is set to 3. Figure 5 shows the coordinates of sentiment words in the 3D space. Each point refers to a sentiment word (Red for positive and blue for negative). As we can see, in the emoticon space, the positive words and the negative words are approximately linearly separable, which is in consistent with our observation that an SVM with a linear kernel achieves the best performance for ESMs in polarity classification.

There still are considerable noises in figure 5. In our experiment, to classify the positive words and the negative words, ESM obtains an accuracy of 0.897 in a five-fold cross validation. After the coordinates are reduced to the 3D space, the accuracy is 0.835.

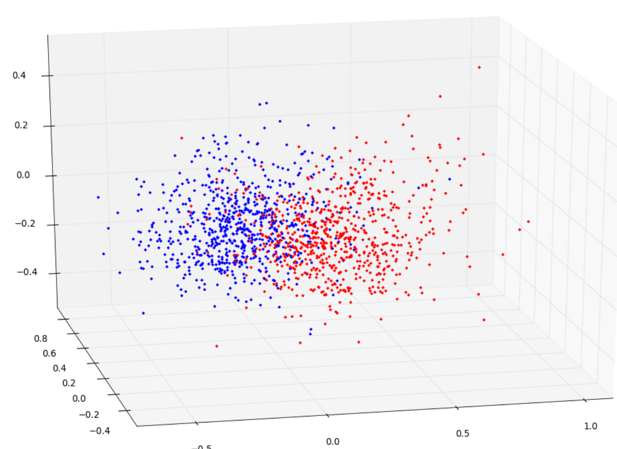


Fig. 5. Sentiment Words in Emoticon Space.

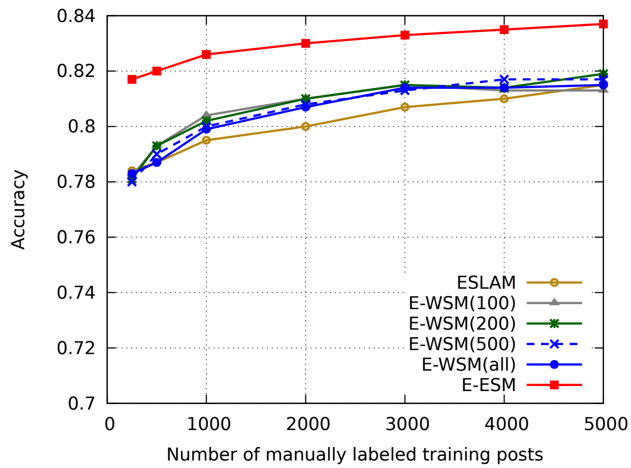


Fig. 6. Comparison between Emoticon Space Model and Word Space Model in polarity classification.

4.8 Comparison to Best Benchmark Results

The original dataset (NLP&CC 2013) was primarily used for an emotion recognition benchmark. To demonstrate the effectiveness of the emoticon space, we use the ESMs to accomplish this benchmark and to compare its results to the best runs in this benchmark. By strictly following the benchmarking procedure, we use the training set in Table 1 for training and validation and the testing set for evaluation. For both the B-ESM and the E-ESM, we first use a subjectivity classifier to identify subjective posts. Next, these posts are classified by a seven-class emotion classifier.

The metrics for this benchmark include: 1) F1 score of the subjective class (the combination of seven emotion types, Subj F1); 2) Micro/Macro average of the F1 score for all emotion types (Micro F1, Macro F1)[1]. All results are illustrated in Table 3. The best runs for the metrics in this benchmark are abbreviated as BEST*. The superiority of the ESMs over BEST* is listed in parentheses. ESM achieves improvements of 8.1% on Subj F1, 24.7% on Micro F1 and 11.8% on Macro F1.

5 Conclusion

In this paper, we propose the emoticon space model (ESM) for microblog sentiment analysis. By differently treating each emoticon and integrating emoticons that do not have clear emotional meanings, ESM effectively leverages emoticon signals and consistently outperforms previous state-of-the-art methods. ESM is flexible to introduce more features of the posts and the E-ESM outperforms the B-ESM. Moreover, ESM obtains a relatively high performance when the size of manually labeled training set is small and can effectively perform lexicon-based polarity classification which does not require manually labeled data.

We also show that emoticons provide richer signals than sentiment words to construct sentiment spaces. The emoticons are irreplaceable in similar models.

Many treasures can be further acquired from

1. http://tcci.ccf.org.cn/conference/2013/dldoc/ev02.pdf

Table 3. Comparison between best runs of NLP&CC 2013 and ESMs

	Subj F1	Micro F1	Macro F1
BEST*	0.729	0.352	0.313
B-ESM	0.782(7.3%)	0.416(18.2%)	0.329(5.1%)
E-ESM	0.788(8.1%)	0.439(24.7%)	0.350(11.8%)

ESM. Currently, post projection and supervised classification are two separated phases and labeled data do not improve the projection phase. In the future, we will investigate how to seamlessly integrate the two phases. Besides, different users may have different ways of using emoticons, which are worth studying.

In this paper, we discover that constructing the emoticon space by emoticons is considerably better than that by sentiment words. However, using representative words along with the emoticons to reinforce the sentiment space may still be helpful. Words generally have more complex senses than emoticons and it is necessary to find better strategies to select a representative set of words. In the future, we will do further research in this direction.

References

[1] Jansen B J, Zhang M, Sobel K, Chowdury A. Twitter power: Tweets as electronic word of mouth. *Journal of the American society for information science and technology*, 2009, 60(11):2169–2188.

[2] Bollen J, Mao H, Zeng X. Twitter mood predicts the stock market. *Journal of Computational Science*, 2011, 2(1):1–8.

[3] Zhao J, Dong L, Wu J, Xu K. MoodLens: An emoticon-based sentiment analysis system for Chinese tweets. In *Proc. KDD*, Aug. 2012, pp. 1528–1531.

[4] Jiang L, Yu M, Zhou M, Liu X, Zhao T. Target-dependent twitter sentiment classification. In *Proc. ACL*, Jun. 2011, pp. 151–160.

[5] Liu K L, Li W J, Guo M. Emoticon smoothed language models for twitter sentiment analysis. In *Proc. AAAI*, Jul. 2012.

[6] Bermingham A, Smeaton A F. Classifying sentiment in microblogs: Is brevity an advantage? In *Proc. the 19th ACM International Conference on Information*

and Knowledge Management, Oct. 2010, pp. 1833–1836.

[7] Kouloumpis E, Wilson T, Moore J. Twitter sentiment analysis: The good the bad and the omg! In *Proc. ICWSM*, Jul. 2011.

[8] Barbosa L, Feng J. Robust sentiment detection on twitter from biased and noisy data. In *Proc. the 23rd International Conference on Computational Linguistics: Posters*, Aug. 2010, pp. 36–44.

[9] Pak A, Paroubek P. Twitter as a corpus for sentiment analysis and opinion mining. In *Proc. LREC*, May 2010.

[10] Go A, Bhayani R, Huang L. Twitter sentiment classification using distant supervision. *Technical report, Stanford*, 2009.

[11] Weichselbraun A, Gindl S, Scharl A. Enriching semantic knowledge bases for opinion mining in big data applications. *Knowledge-Based Systems*, 2014, 69:78–85.

[12] Nielsen F Å. A new ANEW: Evaluation of a word list for sentiment analysis in microblogs. *arXiv preprint arXiv:1103.2903, http://arxiv.org/abs/1103.2903v1, Jun. 2015*, 2011.

[13] Esuli A, Sebastiani F. Sentiwordnet: A publicly available lexical resource for opinion mining. In *Proc. LREC*, volume 6, May 2006, pp. 417–422.

[14] Hu X, Tang J, Gao H, Liu H. Unsupervised sentiment analysis with emotional signals. In *Proc. the 22nd international conference on World Wide Web*, May 2013, pp. 607–618.

[15] Cui A, Zhang M, Liu Y, Ma S. Emotion tokens: Bridging the gap among multilingual twitter sentiment analysis. In *Information Retrieval Technology*, pp. 238–249. Springer, 2011.

[16] Bifet A, Frank E. Sentiment knowledge discovery in twitter streaming data. In *Discovery Science*, Oct. 2010, pp. 1–15.

[17] Davidov D, Tsur O, Rappoport A. Enhanced sentiment learning using twitter hashtags and smileys. In *Proc. the 23rd International Conference on Computational Linguistics: Posters*, Aug. 2010, pp. 241–249.

[18] Bengio Y, Ducharme R, Vincent P, Jauvin C. A neural probabilistic language model. *Journal of Machine Learning Research*, 2003, 3:1137–1155.

[19] Mnih A, Hinton G E. A scalable hierarchical distributed language model. In *Proc.*

Advances in neural information processing systems, Dec. 2008, pp. 1081–1088.

[20] Mikolov T, Sutskever I, Chen K, Corrado G, Dean J. Distributed representations of words and phrases and their compositionality. *arXiv preprint arXiv:1310.4546, http://arxiv.org/abs/1310.4546v1, Jun. 2015*, 2013.

[21] Mikolov T, Chen K, Corrado G, Dean J. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781, http://arxiv.org/abs/1301.3781v3, Jun. 2015*, 2013.

[22] Zhang H P, Yu H K, Xiong D Y, Liu Q. HHMM-based Chinese lexical analyzer ICTCLAS. In *Proc. the second SIGHAN workshop on Chinese language processing-Volume 17*, Jul. 2003, pp. 184–187.

[23] Chang C C, Lin C J. Libsvm: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 2011, 2(3):27.

[24] Zhang W, Liu J, Guo X. Sentiment lexicon for students. *Encyclopedia of China Publishing House (in Chinese)*, 2004.