

Microblog Sentiment Analysis with Emoticon Space Model

Fei Jiang^{1,2,3} (姜 飞), Yi-Qun Liu^{1,2,3,*} (刘奕群), *Senior Member, CCF*, Huan-Bo Luan^{1,2,3} (栾焕博)
Jia-Shen Sun⁴ (孙甲申), Xuan Zhu⁴ (朱 璇), Min Zhang^{1,2,3} (张 敏), *Senior Member, CCF*
and Shao-Ping Ma^{1,2,3} (马少平)

¹State Key Laboratory of Intelligent Technology and Systems, Beijing 100084, China

²Tsinghua National Laboratory for Information Science and Technology, Beijing 100084, China

³Department of Computer Science and Technology, Tsinghua University, Beijing 100084, China

⁴Language Computing Laboratory, Samsung Research & Development Institute of China, Beijing 100028, China

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

Received November 15, 2014; revised July 3, 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 studies 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 meanings. 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 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 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 kind of 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 sen-

timent 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

Regular Paper

Special Section on Social Media Processing

This work was supported by Tsinghua-Samsung Joint Laboratory, the National Basic Research 973 Program of China under Grant No. 2015CB358700, and the National Natural Science Foundation of China under Grant Nos. 61472206, 61073071, and 61303075.

A preliminary version of the paper was published in the Proceedings of SMP 2014.

*Corresponding Author

©2015 Springer Science + Business Media, LLC & Science Press, China

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 one 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 studies use these emoticons as noisy sentiment labels of microblog posts to train classifiers^[3,9-10] 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 that of the emoticons. Thus, noises are introduced. Second, every emoticon has 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 cannot be 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 treat-

ing 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 (e.g., 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 and Smeaton^[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] respectively. Kouloumpis *et al.*^[7] and Barbosa *et al.*^[8] used sentiment lexicons as features to reinforce supervised classifiers respectively.

Supervised methods are effective for microblog sentiment analysis^[4,6]. Birmingham and Smeaton^[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 and Paroubek^[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 and Frank^[16] and Go *et al.*^[10] respectively. 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 methods as noisy supervised methods in the remainder of this paper.

Additionally, noisy supervised methods are adopted to subjectivity classification. Pak and Paroubek^[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 work cannot 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 men-

tioned in Section 1, 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 as 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 $\mathbf{M}_w \in \mathbb{R}^{d \times V}$, where V is the size of the vocabulary and d is the dimension of the representation vectors. Each column of \mathbf{M}_w denotes the representation vector of the corresponding word.

Suppose E emoticons (denoted as (e_1, e_2, \dots, e_E)) are selected to construct the emoticon space. We search for the representation vectors of these emoticons in matrix \mathbf{M}_w and receive a matrix $\mathbf{M}_e \in \mathbb{R}^{d \times E}$. \mathbf{M}_e is a

sub-matrix of M_w . Each column in 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 (1).

$$\text{similarity}(w_i, e_j) = \frac{w_i \cdot e_j}{|w_i||e_j|}, \quad (1)$$

where w_i and 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 (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 C .

Algorithm 1. Calculation of the Coordinates of Words and Emoticons

Require:

- Distributed representation matrix of words and emoticons, M_w
 - Distributed representation matrix of emoticons, M_e
 - 1: **for** each i in $[1 : E]$ **do**
 - 2: **for** each j in $[1 : V]$ **do**
 - 3: $C(i, j) = \text{similarity}(M_w(:, j), M_e(:, i))$
 - 4: **end for**
 - 5: **end for**
-

In Algorithm 1, $C \in \mathbb{R}^{E \times V}$, and each column of 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 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 Subsection 3.1, 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 posts and emoticons, however, cannot 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 \mathbf{p} be the coordinate of the post. Therefore, \mathbf{p} can be computed with (2).

$$\mathbf{p} = \sum_{w_j \in p} C(:, j). \quad (2)$$

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

$$\mathbf{p} = \begin{pmatrix} (\min_{w_j \in p} C(1, j), \sum_{w_j \in p} C(1, j), \max_{w_j \in p} C(1, j)) \\ (\min_{w_j \in p} C(2, j), \sum_{w_j \in p} C(2, j), \max_{w_j \in p} C(2, j)) \\ \vdots \\ (\min_{w_j \in p} C(E, j), \sum_{w_j \in p} C(E, j), \max_{w_j \in p} C(E, j)) \end{pmatrix}. \quad (3)$$

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 Experimental Setup

Experimental studies are performed on a public Chinese microblog benchmark corpus (NLP&CC 2013). This corpus consists of both a training set and a test set^{①~③}. The dataset is composed of 14 thousand microblog posts collected from Sina Weibo and each of them is annotated with neutral or one of the seven

emotion tags: 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, happiness and like are combined together as the positive sentiment class, while sadness, anger, disgust, and fear are combined together as the negative sentiment class^④. In subjectivity classification, neutral is regarded as the objective class and the other seven emotion types are combined together as the subjective class. In emotion classification task, each emotion type serves as one class.

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

| | Neutral | Like | Happiness | Sadness | Disgust | Anger | Surprise | Fear |
|--------------|---------|-------|-----------|---------|---------|-------|----------|------|
| Training set | 1 823 | 597 | 371 | 388 | 425 | 235 | 112 | 49 |
| Test set | 4 925 | 1 525 | 1 106 | 759 | 969 | 405 | 221 | 90 |
| Total | 6 748 | 2 122 | 1 477 | 1 147 | 1 394 | 640 | 333 | 139 |

To evaluate the performance of the proposed ESM model, 100 emoticons^⑤ which are the most frequently used in microblog posts are selected to 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. Similar to Bermingham and Smeaton^[6], we use binary feature vectors for SVM. 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 emoticons mentioned above as noisy sentiment labels. Table 2 shows the five most frequently adopted emoticons for each sentiment or emotion. The positive emoticons and the negative emoticons are adopted in polarity classification and the combination of them is adopted for the subjective class in subjectivity classification. The emoticons for happiness, like, sadness, disgust (the other three emotions are not used as explained later) are adopted in emotion classification. The assumption of Pak and Paroubek^[9] is adopted for the noisy objective label and outperforms the assumption of Liu *et al.*^[5] We implement a noisy MNB (i.e., N-MNB) classifier, which is adopted by most previous work.

Table 2. Emoticons with Clear Emotional Meanings

| Sentiment | Example Emoticons | Amount |
|-----------|-------------------|--------|
| Positive | | 33 |
| Negative | | 19 |
| Happiness | | 9 |
| Like | | 10 |
| Sadness | | 7 |
| Disgust | | 5 |

^①<http://tcci.ccf.org.cn/conference/2013/dldoc/evsam02.zip>, July 2015.

^②<http://tcci.ccf.org.cn/conference/2013/dldoc/evdata02.zip>, July 2015.

^③<http://tcci.ccf.org.cn/conference/2013/dldoc/evans02.zip>, July 2015.

^④We do not use emotion of surprise for polarity classification, as the polarity of this emotion is undetermined in this dataset.

^⑤Except for the results in Fig.2.

- *Combination Methods.* Liu et al.^[5] proposed a state-of-the-art method that effectively integrates the supervised methods and the noisy supervised methods. The 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, E-ESM(*) and 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 Liu et al.^[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, 1 000, 2 000, 3 000, 4 000, and 5 000.

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 Fig.1. From this figure, we can see that B-ESM(*) and 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. ESMs consistently outperform baseline methods for different training dataset sizes. According to the statistics, E-ESM slightly outperforms B-ESM because more features of the posts are introduced. When the training dataset size is larger than 4 000, the performance differences between ESLAM and the supervised methods (S-MNB and S-SVM) become closer. However, a large gap still exists between the baseline methods and ESMs.

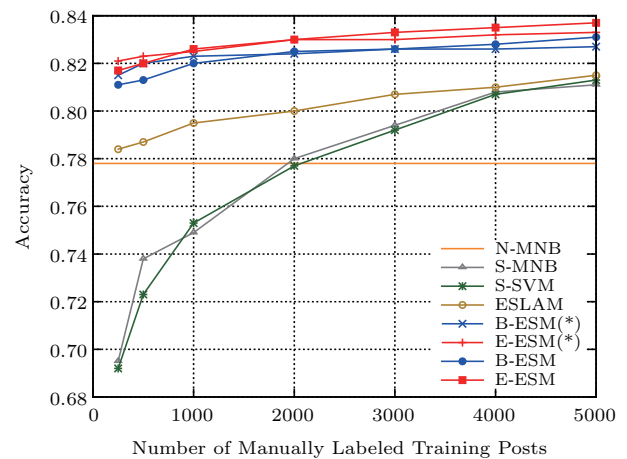


Fig.1. Accuracies for different numbers of manually labeled training posts in polarity classification.

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

We then investigate the performance of B-ESM and 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 ESLAM. The models are trained with 5 000 manually labeled posts. Fig.2 shows the performance of the models as N varies. We can see that with more emoticons adopted, ESMs achieve a better performance and outperform ESLAM. Moreover, ESMs are flexible to leverage much more emoticons at no extra cost except for a few more computations.

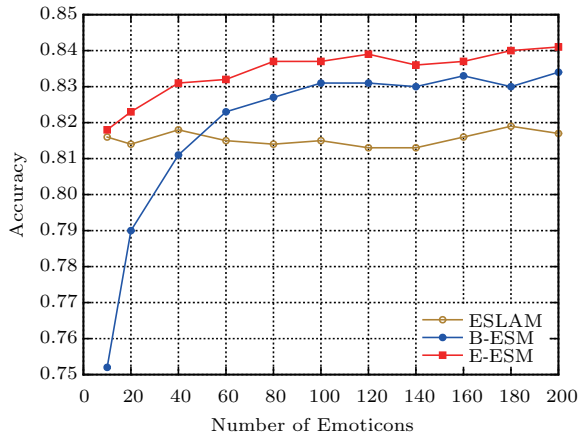


Fig.2. Accuracies of ESMS and ESLAM using different numbers of emoticons in polarity classification.

4.3 Subjectivity Classification

Similar to polarity classification, we randomly sample a balanced dataset of 13496 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 Fig.3.

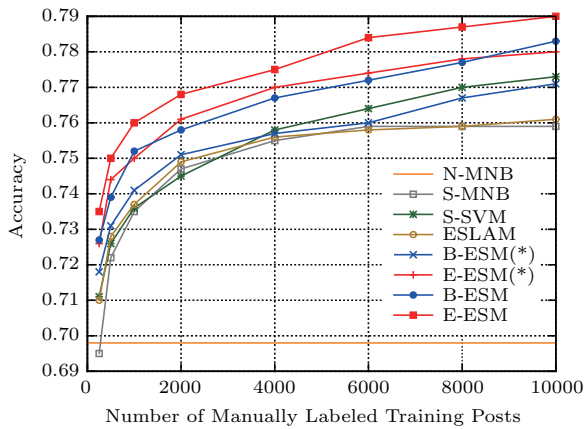


Fig.3. Accuracies for different numbers of manually labeled training posts in subjectivity classification.

For baseline methods, when the training size is larger than 4000, S-SVM outperforms both S-MNB and ESLAM, and even slightly outperforms B-ESM(*). However, the other three ESMS outperform all baseline methods consistently, for different training sizes. All ESMS outperform the noisy supervised methods and ESLAM. Meanwhile, after comparing E-ESM with E-ESM(*) and B-ESM with B-ESM(*), we can see that emoticons that do not have clear emotional meanings help subjectivity classification.

In subjectivity classification, ESMS use more training data to obtain a relatively high performance when

compared with polarity classification. However, to achieve a comparable accuracy, B-ESM and 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 (anger, fear, and neutral) are relatively small. Similar to polarity classification, a balanced set of 4588 posts is sampled and contains 1147 posts for each type. The size of the manually labeled training set varies as 400, 800, 1600, and 3200.

Fig.4 shows the results for this task. The results are similar to polarity classification. E-ESM outperforms E-ESM(*) when the training size is no less than 800. ESMS outperform all baseline methods for different training sizes and they are robust and achieve high performances when the training size is relatively small. For example, 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 3200.

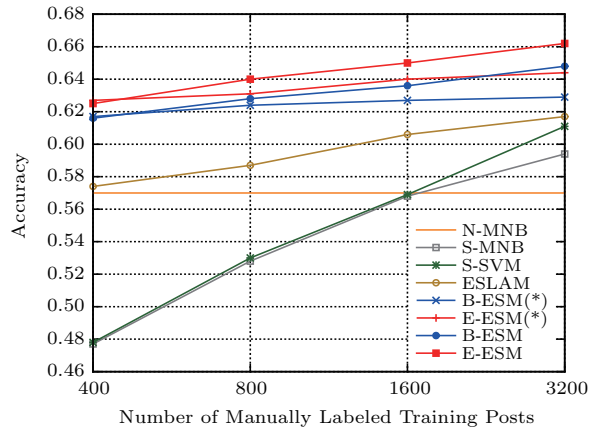


Fig.4. Accuracies for different numbers of manually labeled training posts in emotion classification.

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 6640 posts above as the test data for B-ESM(*). This procedure is performed for fixed parameters in ten rounds. B-ESM(*)

achieves an accuracy of 0.803, which slightly outperforms ESLAM with a training size of 2000. 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 with 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 from [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 the 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.

Fig.5 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.

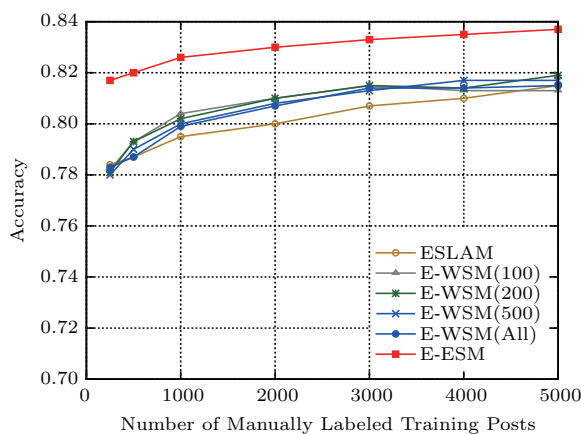


Fig.5. Comparison between the Emoticon Space model and the Word Space model in polarity classification.

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, and so on.

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. Fig.6 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 consistent with our observation that an SVM with a linear kernel achieves the best performance for ESMs in polarity classification.

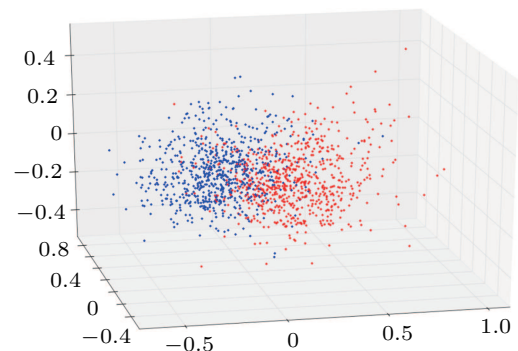


Fig.6. Sentiment words in emoticon space.

There still are considerable noises in Fig.6. In our experiment, to classify the positive words and the negative words, ESM obtains an accuracy of 0.897 in a 5-fold cross validation. After the coordinates are reduced to the 3D space, the accuracy is 0.835.

4.8 Comparison with 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 with 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 B-ESM

and E-ESM, we first use a subjectivity classifier to identify subjective posts. Next, these posts are classified by a 7-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 of the seven emotion types (micro $F1$, macro $F1$)^⑥. 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 $F1$.

Table 3. Comparison of Best Runs Between 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%) |

5 Conclusions

In this paper, we proposed 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 E-ESM outperforms 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 showed 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 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 discovered that constructing the emoticon space by emoticons is considerably better

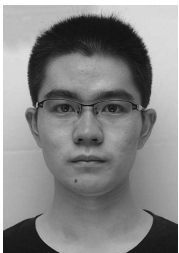
than that by sentiment words. However, using representative words along with the emoticons to reinforce the sentiment space may still be helpful. Words generally have more complex senses than emoticons and it is necessary to find better strategies to select a representative set of words. In the future,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

References

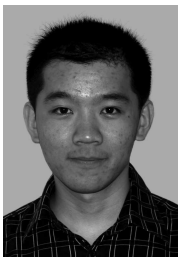
- [1] Jansen B J, Zhang M, Sobel K, Chowdury A. Twitter power: Tweets as electronic word of mouth. *Journal of the American Society for Information Science and Technology*, 2009, 60(11): 2169-2188.
- [2] Bollen J, Mao H, Zeng X. Twitter mood predicts the stock market. *Journal of Computational Science*, 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. the 18th KDD*, Aug. 2012, pp.1528-1531.
- [4] Jiang L, Yu M, Zhou M, Liu X, Zhao T. Target-dependent Twitter sentiment classification. In *Proc. the 49th 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. the 26th AAAI*, Jul. 2012.
- [6] Birmingham 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. the 5th 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 University, 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:1103.2903*, 2011. <http://arxiv.org/abs/1103.2903v1>, Jun. 2015.
- [13] Esuli A, Sebastiani F. SENTIWORDNET: A publicly available lexical resource for opinion mining. In *Proc. the 5th LREC*, 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 *Proc. the 7th AIRS*, Dec. 2011, pp.238-249.
- [16] Bifet A, Frank E. Sentiment knowledge discovery in Twitter streaming data. In *Proc. the 13th DS*, Oct. 2010, pp.1-15.

⑥ <http://tcci.ccf.org.cn/conference/2013/dldoc/ev02.pdf>, July 2015.

- [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. the 22nd 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:1310.4546*, 2013. <http://arxiv.org/abs/1310.4546v1>, Jun. 2015.
- [21] Mikolov T, Chen K, Corrado G, Dean J. Efficient estimation of word representations in vector space. *arXiv:1301.3781*, 2013. <http://arxiv.org/abs/1301.3781v3>, Jun. 2015.
- [22] Zhang H P, Yu H K, Xiong D Y, Liu Q. HHMM-based Chinese lexical analyzer ICTCLAS. In *Proc. the 2nd 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): 27:1-27:27.
- [24] Zhang W, Liu J, Guo X. Sentiment Lexicon for Students. Encyclopedia of China Publishing House, 2004. (in Chinese)



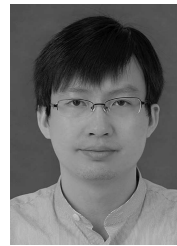
Fei Jiang is a master candidate in Department of Computer Science and Technology, Tsinghua University, Beijing. He received his B.S. degree in computer science from Tsinghua University in 2013. His research interests include information retrieval and natural language processing.



Yi-Qun Liu is an associate professor in Tsinghua University, Beijing. He received his B.S and Ph.D. degrees in computer science from Tsinghua University in 2003 and 2007, respectively. His research interests include Web search, user behavior analysis and performance evaluation.



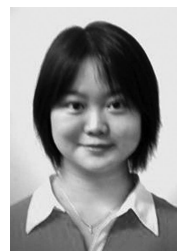
Huan-Bo Luan is the deputy executive director of NExT Search Center at both Tsinghua University and National University of Singapore. He received his B.S. degree from Shandong University, Jinan, in 2003, and Ph.D. degree from Institute of Computing Technology, Chinese Academy of Sciences, Beijing, in 2008, both in computer science. His research interests include multimedia information retrieval, social media and big data analysis.



Jia-Shen Sun is a senior engineer in Language Computing Laboratory, Samsung Research & Development Institute of China, Beijing. He received his B.S. degree in information engineering in 2006, and Ph.D. degree in computer science in 2013, both from Beijing University of Posts and Telecommunications, Beijing. His research interests are natural language processing, text mining, and market intelligence.



Xuan Zhu is a principal researcher in Language Computing Laboratory, Samsung Research & Development Institute of China, Beijing. She received her B.S. degree from Beihang University, Beijing, in 1998, and M.S. and Ph.D. degrees from Tsinghua University, Beijing, in 2000 and 2003 respectively, all in electronic engineering. Her research interests are natural language processing, data analytics, speech recognition, speech synthesis, and market intelligence.



Min Zhang is an associate professor in Tsinghua University, Beijing. She received her B.S and Ph.D. degrees in computer science from Tsinghua University in 1999 and 2003, respectively. Her research interests include Web search and recommendation, natural language processing, and machine learning.



Shao-Ping Ma is a professor in Tsinghua University in Beijing. He received his B.S, M.S. and Ph.D. degrees in computer science from Tsinghua University in 1982, 1984 and 1997, respectively. His research interests include Web search, knowledge engineering, and artificial intelligence.