Microblog Sentiment Analysis Based on Cross-media Bag-of-words Model

Min Wang
Dept. of Cognitive Science
Xiamen University
Xiamen, Fujian, China
wxm_xmu@163.com

Donglin Cao*, Lingxiao Li
Dept. of Cognitive Science
Xiamen University
Xiamen, Fujian, China
another@xmu.edu.cn,
llx1990rl@sina.com

Shaozi Li, Rongrong Ji
Dept. of Cognitive Science
Xiamen University
Xiamen, Fujian, China
{szlig, rrji}@xmu.edu.cn

ABSTRACT
Sentiment analysis on social networks has attracted ever increasing attention in recent years. To this end, most existing methods mainly focus on analyzing the textual information. This has faced huge difficulty as nowadays users are more likely to express their feelings in a hybrid manner not only with texts, but also with images. It is therefore essential to take images into account. In this paper, we propose a novel Cross-media Bag-of-words Model (CBM) for Microblog sentiment analysis. In this model, we represent the text and image of a Weibo tweet as a unified Bag-of-words representation. Based on this model, we use Logistic Regression to classify the Microblog sentiment. It performs well in the sentiment classification task since it doesn't require the conditional dependence assumption. We also use SVM and Naïve Bayes to make a comparison. Experiments on 5,000 Microblog messages demonstrate that our CBM model performs better than text-based methods. The sentiment classification accuracy on Microblog messages of our model is 80%, improved by 4% than the text-based methods.

Categories and Subject Descriptors
H.2.4 [**Database Management**]: Systems – *Multimedia databases;* H.2.8 [**Database Management**]: Database Applications – *Data mining*

Keywords
Sentiment analysis, Microblog, Cross-media, Bag-of-words, Logistic Regression.

1. INTRODUCTION
Nowadays have witnessed a rapid development of social networks such as Twitter, Facebook, Sina Weibo and Tencent Weibo. Taking Twitter as example, it's reported that the number of Twitter registered users is over 500 million by July 1, 2012. As another example Sina Weibo has attracted more than 500 million users who post around 100 million messages a day. Despite its ever-increasing number, the value and strength of Microblog has become higher and stronger with the development of itself.

ICIMCS'14, July 10–12, 2014, Xiamen, Fujian, China.
Copyright 2014 ACM 978-1-4503-2810-4/14/07 …$15.00.

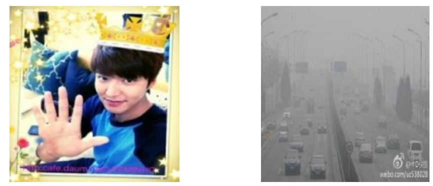


Figure 1: Microblog messages with images. The actor in play <The heirs> on the left, with text "he posted a self-taken photo, every one begins to play". The misty picture on the right with text "The fog is coming with steps like a kitten".

Mining knowledge from such massive amount of data is becoming a research hotspot very recently. Among them, mining the sentiment from Microblog is ever more focused, which has a wide variety of applications in commercial recommendation, event monitoring, and social network analytics.

One emerging trend in the data constitution of Microblog is its ever increasing proportion of visual content. As we can see, an increasing number of Microblog messages are appended with images in addition to texts. This is especially the case when these contents are sent via mobile platform e.g. cellphones, in which scenario is less convenient for the users to type long paragraph, while very convenient to take pictures. For example, Figure 1 contains two Weibo tweets, the left one is positive and the right one is negative. Apparently, the message sentiment can be seen in images, not texts. This means that the image is really helpful for the Microblog sentiment analysis. To understand the visual sentiment delivered, Borth et al. proposed a visual sentiment ontology together with a so-called sentibank detector [1]. However their method mainly focuses on the image analysis, where cross-media analysis is simply achieved by fusing the scores of both text-based and image-based sentiment prediction.

In this paper, we argue that the combination of visual and textual sentiment predictions should be on the basis of visual level, i.e., a cross-media feature representation, as to being independent for the subsequent classification steps. For example, given such a representation at hand, several cutting-edge classifiers can be simply plug in for testing the overall classification performance. There are other social media researches combining texts and images like travel recommendation and social image search [7, 8].

We propose a novel cross-media bag-of-words model for Microblog sentiment analysis. In our model, the text and image are all represented as a bag of words, namely the feature vector of

* Corresponding author：Donglin Cao

a message. The bag-of-words representation takes texts and images as an entirety, in this way text and image can be managed by a uniform approach, regardless of the difference of their low-level features. Our classification model is trained using Logistic Regression approach with the labeled data considering it can not only be used for probability prediction and classification, but also not need to meet conditional dependence assumption. We also use SVM, Naïve Bayes method to make comparisons. Finally our model is proved to be better than the state-of–art text based method with accuracy improved by 4%.

Overall our main contributions are: Firstly, we take images into consideration in our Microblog sentiment analysis task. Secondly, we present a novel cross-media bag-of-words model for sentiment classification combining texts and images. In this model, the text and image are represented as an entirety using bag-of-words method. The word feature in our model is specially selected considering the unique attributes of Microblog messages. We can get a universal function to deal with the image and text. Lastly, Logistic Regression is adopted to classify the messages, which is proved to have good performance in our experiments.

The rest of this paper is organized as follows. We first discuss related work in section 2, then showing details of our model for sentiment classification in section 3. Section 4 is the experiment results. Conclusion and future work are described in section5.

2. RELATED WORK

Traditional sentiment analysis always focuses on the text and has achieved much progress. By construct, works on image sentiment detection are much fewer. We will discuss approaches of text sentiment and image sentiment below. Besides, with respect to the traditional sentiment analysis on blogs and product reviews, sentiment analysis of Microblog has attracted much attention.

2.1 Text Sentiment

There are mainly two approaches on text sentiment analysis, one is dictionary based method, and the other is machine learning method.

Dictionary based method, also called supervised learning method, always depending on the pre-existing sentiment dictionaries and sentiment phrases. A typical method was proposed by Turney (2002) [2], in which sentimental phrases are first selected from the reviews according to predefined part-of-speech patterns. Then the semantic orientation score of the phrase is calculated by point-wise mutual information measures. The review is classified according to the average sentiment orientation of the sentiment phrases in it. Besides, a widely used approach is using a dictionary of sentiment words and phrases. First is to obtain the orientation of the words or phrases according to the dictionary, and then compute the weighted sum of the orientation. [3, 4, 5]

With regard to machine learning method, Pang [6] applied the machine learning approaches to the chapter-level sentiment classification task for the first time. They take n-gram and parts of speech as features, classifying texts with Naïve Bayes, MaxEnt and SVM methods. Experiments imply that the unigram feature combining with Naïve Bayes or SVM approach has better performance than other features and methods.

2.2 Image Sentiment

There are not so many researches on image sentiment analysis as text. D Borth et al. propose a novel approach to construct visual sentiment ontology and detectors, then applying it to sentiment or emotion classification tasks [1]. A new mid-level representation called Adjective Noun Pairs is used in their paper. Researches similar to sentiment analysis are aesthetics [9, 10], affect or emotions of images [11, 12, 13]. We adopt D Borth's method to get features of Microblog images.

2.3 Sentiment Analysis of Microblog

Microblog is a Twitter like website. There are an increasing number of researches on Twitter sentiment analysis due to the rapidly developing social networks. Go et al.2009; Barbosa and Feng, 2010; Davidiv et al. 2010 classify the tweets with machine learning methods [14~16]. Go et al. classify the messages as either positive or negative using tweets labeled with distant supervised learning method. Jiang et al. incorporate target-dependent features and take related tweets into consideration [17]. However, researches on Chinese Microblog messages are not as many and well as on Twitter. One important reason is the complexity of the Chinese sentence representation. To our best knowledge, there are no studies on sentiment analysis of Chinese Microblog messages combining texts and images. D Borth's method mainly focuses on the establishment of visual sentiment ontology and performs sentiment prediction with simple weight fusion of images and texts. Our model is different because of taking images and texts as an entirety using special bag-of-words method and adopting a universal classifier.

3. METHODS

3.1 Problem Description

Traditional researches on sentiment analysis always focus on text, leaving images out of consideration. The classification task can be expressed by function $y_t = f_t(T)$, where T is the text and $\mathrm{y_t}$ is the prediction result by function f_t. T can be represented as a bag of words like $T = \{\mathrm{w_1}, \dots, \mathrm{w_j}, \dots, \mathrm{w_m}\}$. Function$f_t$ could be learned by machine learning methods such as SVM, Naïve Bayes and so on. We choose Logistic Regression, SVM and Naïve Bayes to be the text sentiment classification functions, which serve as the baseline comparing with our CBM model containing images.

In this paper we are interested in analyzing Microblog sentiment combining text and image. The key point is to learn a classifier $y_{tp} = f_{tp}(x)$ from a set of training examples, where $x = \{T, I\}$. An easy way to get the sentiment value of a message containing text and image is to calculate the value of text and image respectively, then taking weighted summation method to get the final result as follows:

$$\mathrm{y} = \lambda f_t(T) + (1-\lambda) f_p(I) \tag{1}$$

The expression $f_p(I)$ gives the sentiment classification result of image I. Equation (1) combines the result of text and image by a weight coefficient λ, which can be assigned manually or automatically learned from the labeled data $D = \{(\mathrm{x_1}, \mathrm{y_1}), \dots, (\mathrm{x_i}, \mathrm{y_i}), \dots, (\mathrm{x_n}, \mathrm{y_n})\}$, where $\mathrm{x_i} = \{\mathrm{T_i}, \mathrm{I_i}\}$ is a message and $\mathrm{y_i}$ is the category label of it. The value of $\mathrm{y_i}$can be 1, 0 or-1, which denotes positive, neutral and negative respectively.

Nevertheless, there are some disadvantages in equation (1). Firstly, this equation can just get external links of the text and the image. As a matter of fact, there is some internal relation between them. For example, a message about a typhoon may have "house damage" in the text and the corresponding visual content in the picture. Secondly, two classifiers and a coefficient need to be trained in equation (1). As a result, we propose our CBM (Cross-media Bag-of-words Model) model. In this model, texts and images are represented as an entirety using bag-of-words method. In this way, texts and images can be classified by a universal classifier and some internal relations which are useful for the sentiment analysis may be found. We will give a detail explanation in section 3.2.

3.2 CBM Model

We use bag-of-words method to represent a Microblog message as $x = \{w_1, \dots, w_i, \dots, w_{m,} p_1, \dots, p_j, \dots, p_d\}$ in our CBM model, which consists of text words $w_1 \sim w_m$ and image words $p_1 \sim p_d$, thus images and texts can be considered as an entirety. The advantage of this representation is that we can use a universal classifier to deal with the messages, ignoring it is text or image. Besides, this representation can get better results than the weighted summation approach because of the internal relations of texts and images.

As we all know, the low-level feature of images and texts are totally different. The text can actually be expressed as word vectors. However, the image features are always color histograms, textures or shapes, which can also be transferred to bag-of-words models by clustering. However, the dimension of text words and image words may differ a lot. We choose special word representation for texts and images in our model, which makes the dimension of image feature and text feature almost equals. We will talk about bag of text words and bag of image words in section *3.2.1* and *3.2.2*. The final features of a message are composed of text words and image words.

To achieve a better performance of classification, the Logistic Regression is adopted in our CBM model. The reason of choosing this method is that the Logistic Regression doesn't need to meet conditional independence assumption between the features, unlike Naïve Bayes. Experiments show that it achieves a better performance than the Bayes and SVM method. We will talk about it in section *3.2.3*

3.2.1 Bag of Text Words

For traditional text classification or English sentiment analysis, the features of text are usually unigram or bigram [6]. Nevertheless, it is hard for a sentence in Chinese with this representation because of two reasons. First, the Chinese sentence needs word segmentation, thus the total number of all the words within our data is really giant. We investigate our data set with 5000 messages, and finally get 10833 different segmented words. Second, the length of Microblog message is too short with less than 140 words. As a result, features could be extremely sparse. Therefore, we consider choosing some simple but discriminating features for text representation, which shows in table 1.

According to table 1, we can represent text T_i as a bag of words like $T_i = \{w_{i,1}, \dots, w_{i,j}, \dots, w_{i,m}\}$, $w_{i,1} \in [0, +\infty)$ represents the number of positive words in text, the other $w_{i,j}$ is similarly defined in table 1. The dimension of feature T_i in this paper is ten. A preprocessing step before feature extraction is the text segmentation, which is performed by ICTCLAS [18]. The dictionary we use is the incorporation of HowNet and NTUSD. We take text-based method as a baseline in our experiment, in which Naive Bayes, SVM and Logistic Regression are applied to do the classification task with feature $T_i$according to B. Pang [6].

3.2.2 Bag of Image Words

There are an increasing number of researches on image sentiment analysis. D Borth and R ji' presents us a new approach. They first retrieve images and videos from Flickr and YouTube to extract concurrent tags for 24 basic emotions. Then adjectives and nouns in these tags are used to construct Adjective Noun Pairs (ANP) such as "happy baby" and "dark clouds". Each ANP has a sentiment value between -2 to 2. There are more than 3000 ANP concepts left after processing, which composes a comprehensive ontology. Finally a detector library called SentiBank is established, including1200 ANP detectors trained from Flickr images with reasonable performance. We apply their method to extract image sentiment features.

The 1200 detectors in SentiBank are used to obtain a mid-level concept representation of the images. The detecting result of an image is a 1200 dimensional probability vector corresponding to 1200 ANPs. As we have a brief short feature of a text, it's essential to simplify the image feature too. We just take top N responding ANP detectors, then calculating the number of positive and negative ANPs. N is endowed with different values thus the appropriate N could be recognized by the classification approach. In this paper we choose five values 10, 20, 50, 100, 200 for N. For each N, we count the number of positive ANPs and negative ANPs. Then the image feature can be represented as $I_i = \{p_{i,1}, \dots, p_{i,k}, \dots, p_{i,d}\}$, where d=10 in this paper.

Table 1: Text Features

No	Feature type	Feature Content	Description
1	Sentiment dictionary	The number of positive sentiment and negative sentiment words (reversing the sentiment of a word when it has a privative).	The number of positive and negative words in the dictionary we use is 8540 and 7754 respectively.
2	Part of speech	The number of nouns, adjectives and verbs.	The message is segmented and POS tagged in advance.
3	Punctuation	The number of exclamation marks and question marks and full stops.	The number of full stops equals one when over two continuous stops occurring, otherwise equals zero.
4	Link	Whether there is a link.	The link begins with http:
5	Negative words	The number of privative words.	

Table 2: Image Features

No.	Feature	Feature Content	Description
1	Top N responding ANP detectors	The number of top N positive ANPs and negative ANPs.	The values of N can be 10, 20, 50, 100, 200 or other positive integers less than 1200.

3.2.3 Microblog Sentiment Classification by Logistic Regression

As mentioned above, text words and image words are prepared. For the messages with images and texts, we can represent it as an entire bag of words:

$$x_i = \{T_i, I_i\} = \{w_{i,1}, \dots, w_{i,j}, \dots, w_{i,m}, p_{i,1}, \dots, p_{i,k}, \dots, p_{i,d}\}$$

Then a classification function $y_{tp} = f_{tp}(x)$ is applied to perform the sentiment prediction. In this paper Logistic Regression is employed to do the task. Logistic Regression is a type of probabilistic statistical classification model. The probability of class k for a given instance x is calculated as follows:

$$\Pr(C=k \mid X=x) = \frac{\exp(\beta_{k0} + \beta_k^T x)}{1 + \sum_{c=1}^{K-1} \exp(\beta_{c0} + \beta_c^T x)} \qquad (2)$$

In equation (2), $k = 1, \ldots, \mathrm{K}-1$

$$\Pr(C=\mathrm{K} \mid X=x) = \frac{1}{1 + \sum_{c=1}^{K-1} \exp(\beta_{c0} + \beta_c^T x)} \qquad (3)$$

The sum of all class probability for instance x is 1:

$$\sum_{c=1}^{K} \Pr(C=c \mid X=x) = 1 \qquad (4)$$

The category $c*$ is assigned to x, where

$$c* = \arg mac_c \Pr(C=c \mid x) \qquad (5)$$

The class label K in this paper equals 2 or 3 (1 is positive, 2 is negative, 3 is neutral). The weight vector β in equation (2) can be found by Maximum Likelihood Estimation from the training examples.

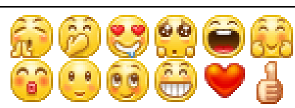
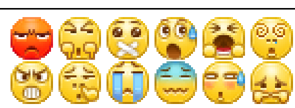
As Pang el al. [6] has proved that SVM and Na ïve Bayes perform well on the sentiment classification, so we adopt these two methods to be the comparison with the same features.

4. EXPERIMENT RESULTS

4.1 Dataset

Our experiment mainly focuses on the data of Sina Weibo. We choose top ten hot topics from the topic list of Sina Weibo, and then extract messages accompanied with images. A set of 5,000 messages is established (2900 positives, 1200 negatives and 900 neutrals). This data set spreads in various fields, like typhoon and haze in social, iOS7 and Meizu MX3 in product, stars and films. We obtain the labeled data by two steps. Firstly, the messages with emotions are picked out. Afterwards, we divide these messages into two classes according to the sentiment of emotions showed in table 3. For the neutral class, we manually label the messages without emotions.

Table 3: Typical emotions with obvious sentiment in Sina Weibo

Positive Emotions	
Negative Emotions	

4.2 Results

Our approach is evaluated in Two-Class (positive and negative) and Three-Class (positive, negative and neutral) way. Besides, we are interested in comparing the results of text-based method (text feature with Na ïve, SVM and Logistic Regression approaches) and cross-media method (cross-media feature with Na ïve, SVM and CBM approaches). The experiment is performed in WEKA with 10 fold cross-validation. Results are shown in table 4, table 5 and figure2.

From table 4 and table 5, we can see that our CBM model get a 4% increase for the Two-Class evaluation compared to text-based method. It also exceeds the SVM and Na ïve Bayes methods with cross-media features for the Two-Class evaluation. From figure 2 we can see the superiority of our CBM model clearly.

Table 6 shows some test examples of our experiment. The text of first two messages seems negative or neutral, but the whole message is positive. Therefore, the text-based method can't get the right answer but our cross-media approach can. The Last two messages are misclassified because the texts use the ironic way to express their feelings and the images have no obvious sentiment.

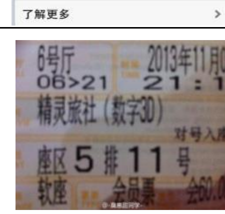
Table 4: Accuracy of text-based method

	Na ïve Bayes	SVM	Logistic Regression
Two-Class	0.72	0.75	0.76
Three-Class	0.63	0.61	0.65

Table 5: Accuracy of cross-media method

	Na ïve Bayes	SVM	CBM
Two-Class	0.73	0.76	0.80
Three-Class	0.64	0.63	0.66

Table 6: Some examples of sentiment classification using our model compared to text-based method

Text	Image	Results
#Yuchun Li's WhyMe concert# 		Ground truth:1 Text-based: -1 CBM:1
#Yundi Li's piano dream # Have a change.		Ground truth:1 Text-based: 0 CBM:1
#ios7#why it takes109 hours to download...The great NTU wireless, you win.		Ground truth:-1 Text-based: 1 CBM:1
It was so late because of watching the movie #Hotel Transylvania#. It is so touching and I cried to tears.		Ground truth:1 Text-based:-1 CBM:-1

For the Results column in table 6, Ground truth is the manually annotated category of the message (-1 is negative, 0 is neutral and 1 is positive). Text-based is the result of the text-based method. CBM represents the result of our cross-media bag-of-words model.

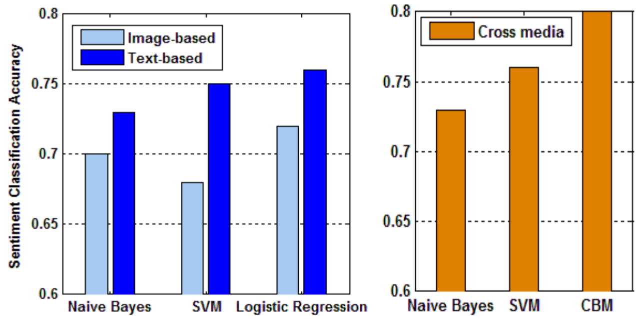


Figure 2: Two-class Sentiment classification with different approaches. The left histogram is the result of image-based and text-based features with Na ïve Bayes, SVM and Logistic Regression approaches. The right one shows result of cross media features with Na ïve Bayes, SVM approaches and our CBM model.

From figure 2, we can see that Logistic Regression performs better than Na ïve and SVM for both image-based and text-based features. The cross media feature with any approach is equal or better than text-based methods, which demonstrates the superiority of combining texts and images. Our CBM model gets the highest accuracy of all the features and approaches.

5. CONCLUSION

To solve the problem of cross-media sentiment analysis problem, we propose a novel approach for Microblog sentiment analysis based on cross-media bag-of-words model (CBM). Previous researches always focus on the text, neglecting the effect of images. Considering more and more people express their feelings through images in Microblog, thus we take images into account in our model. There are three advantages of our method. First, the sentiment of the messages is analyzed by combining the images and texts. Second, this model gives a unified representation of texts and images for cross-media sentiment analysis through bag-of-words method. Finally, we use Logistic Regression to relax conditional independence assumption. Experiment results illustrate the effectiveness of our model, with classification accuracy 4% higher than the text-based method.

There are still some problems needed to be solved. Firstly, we use emotions to label the training data. As a result, emotions couldn't be included in as a feature in the classification step. Secondly, we can't manage gif images or videos. Researches on video sentiment presentation can give us some hints [19]. It's helpful if we get the sentiment of gif images and videos. Finally, many images contain massive texts in them, therefore recognizing the texts in images can be favorable too.

6. ACKNOWLEDGMENTS

This work is supported by the Nature Science Foundation of China (No.61202143 and No. 61305061), the Fundamental Research Funds for the Central Universities (No.2013121026 and No.2011121052), and the National Key Technology Research and Development Program of the Ministry of Science and Technology of China under Grant No. 2013BAB06B04.

7. REFERENCES

[1] D. Borth, R. Ji. Large-scale visual sentiment ontology and detectors using adjective noun pairs. Proc. of the 21st ACM international conference on Multimedia, 2013, 223-232.

[2] Turney P. Thumbs up or thumbs down? : Semantic orientation applied to unsupervised classification of reviews. Proc. of the ACL. Morristown: ACL, 2002, 417-424.

[3] Kim SM, Hovy E. Automatic detection of opinion bearing words and sentences. In: Carbonell JG, Siekmann J, eds. Proc. of the IJCNLP, 2005. Morristown: ACL, 61-66.

[4] Yu H, Hatzivassiloglou V. Towards answering opinion questions: separating facts from opinions and identifying the polarity of opinion sentences. Proc. of the EMNLP, 2003. Morristown: ACL, 129-136.

[5] Hu MQ, Liu B. Mining and summarizing customer reviews. In: Kohavi R, ed. Proc. of the KDD, 2004. New York: ACM Press, 168-177.

[6] B. Pang, L. Lee el al. Thumbs up?: sentiment classification using machine learning techniques. In Proc. of the ACL-02 conference on Empirical methods in natural language processing-Volume 10, 2002, 79-86.

[7] Y. Gao, J. Tang et al. W2Go: A Travel Guidance System by Automatic Landmark Ranking. ACM MM, 2010, 123-132.

[8] Y. Gao, M. Wang et al. Visual-Textual Joint Relevance Learning for Tag-Based Social Image Search. IEEE Transactions on Image Processing, 2013, 363- 376.

[9] R. Datta, D. Joshi, J. Li, and J. Wang. Studying Aesthetics in Photographic Images using a Computational Approach. ECCV, 2006, 288-301.

[10] L. Marchesotti, F. Perronnin, D. Larlus, and G. Csurka. Assessing the Aesthetic Quality of Photographs using Generic Image Descriptors. ICCV, 2011, 1784-1791.

[11] J. Jia, S. Wu, X. Wang, P. Hu, L. Cai, and J. Tang. Can we understand van Gogh's Mood?: Learning to infer A_ects from Images in Social Networks. ACM MM, 2012, 857-860.

[12] J. Machajdik and A. Hanbury. Affective Image Classification using Features inspired by Psychology and Art Theory. ACM Multimedia, 2010, 83-92.

[13] V. Yanulevskaya et al. In the Eye of the Beholder: Employing Statistical Analysis and Eye Tracking for Analyzing Abstract Paintings. ACM MM, 2012, 349-358.

[14] Luciano Barbosa and Junlan Feng. Robust Sentiment Detection on Twitter from Biased and Noisy Data. Coling2010: Poster Volume, 36-44.

[15] Dmitry Davidiv, Oren Tsur and Ari Rappoport. Enhanced Sentiment Learning Using Twitter Hashtags and Smileys. Coling2010: Poster Volume, 23-27.

[16] Alec Go, Richa Bhayani, Lei Huang. Twitter Sentiment Classification using Distant Supervision.2009.

[17] Long Jiang and Mo Yu. Target-dependent Twitter Sentiment Classification. In ACL, 2011, 151-160.

[18] Huaping Zhang el al. HHMM-based Chinese lexical analyzer ICTCLAS. In Proceedings of the second SIGHAN workshop affiliated with 41th ACL, 2003, 184-187.

[19] S. Zhao, H. Yao el al. Flexible presentation of videos based on affective content analysis. Proc. of the 19th International Conference on Multimedia Modeling, 2013, 368–379.