# Emotional Tendency Identification for Micro-blog Topics Based on Multiple Characteristics

**Quanchao Liu**　　　　**Chong Feng**　　　　**Heyan Huang**

Department of Computer Science and Technology

Beijing Institute of Technology

Beijing, China

{liuquanchao, fengchong, hhy63} @bit.edu.cn

## Abstract

Public opinion analysis for micro-blog post is a new trend, and wherein emotional tendency analysis on micro-blog topic is a hot spot in the sentiment analysis. According to the characteristics of contents and the various relations of Chinese micro-blog post, we construct the dictionaries of sentiment words, internet slang and emoticons respectively, and then implement the sentiment analysis algorithms based on phrase path and the multiple characteristics for emotional tendency of micro-blog topics. Using micro-blogs' forwarding, commentaries, sharing and so on, We take a future step to optimize the algorithm based on the multiple characteristics. According to the experimental results, our approach greatly improves the performance of emotional tendency identification on micro-blog topic.

## 1 Introduction

As the fast development of New media, the internet gradually advocates open architecture philosophy that users participate in actively, and has been developed from a simple "reading webpage" to "writing webpage", "building webpage together" for users. So a huge user generated content (UGC) has been developed, especially the emergence of micro-blog post. According to the reports[1], users applying micro-

---

[1] The Eleventh China Network Media Forum

blog have exceeded 300,000,000. As a result of releasing diversely and writing randomly, micro-blog post is more and more popular in China, and it has brought a tremendous effect on network opinions and human society.

Micro-blog contents formed on internet express people's various emotion and sentiment, such as joy, anger, grief, praise, criticism and so on. More and more people like to use micro-blog post to share their views or experience, which makes people's opinion information expanded rapidly. So it is very difficult to rely on the artificial method to deal with the micro-blog information's collection and processing, there is an urgent need to help user analyse the massive information using computer.

Chinese micro-blog post has multiple characteristics as following:

1) Due to the short text message, micro-blog post has terms' sparsity. Feature extraction based on terms is not suitable for micro-blog post.

2) There exist many homophonic words, abbreviated words, internet slang in micro-blog post, such as "杯具" standing for "悲剧", "亲" standing for "亲爱的", "3Q" standing for "谢谢", or using emoticons to express emotion and so on.

3) Many popular new words made by network events would appear in micro-blog post. Taking "皮鞋很忙" for example, it appeared because the news report that the materials for producing shoes are processed into edible gelatin.

4) There are a variety of relations between micro-blogs. It is very convenient to forward, comment and share micro-blog post.

According to the above characteristics of micro-blog post, we analyse emotional tendency on micro-blog topic, and obtain a set of feasible and effective emotional tendency identification algorithm on micro-blog topic.

The remainder of this paper is structured as follows. In section 2, we briefly summarize related work. Section 3 gives an overview of data construction, including emotional symbol library, emotional dictionary and network slang dictionary. In order to improve the rate of target coverage, the target extended algorithm is described in section 4. We design the different emotional tendency identification algorithms on micro-blog topic in sections 5 and 6 respectively. In section 7, sentiment optimization algorithm is described. Experimental results are reported in section 8 and section 9 concludes our work.

## 2 Related Work

In the research domain of sentiment analysis, emotional tendency for twitter has been concerned for some time, such as Tweetfeel[1], Twendz[2], Twitter Sentiment[3]. In previous related work, (Go et al., 2009) use distant learning to acquire sentiment data. They use tweets ending in positive emoticons like ":)" as positive and negative emoticons like ":(" as negative. They build models using Naives Bayes(NB), MaxEnt(ME) and Support Vector Machines(SVM), and they report SVM outperforms other classifiers. In terms of feature space, they try a Unigram, Bigram model in conjunction with parts-of-speech (POS) features. They note that the unigram model outperforms all other models. However, the unigram model isn't suitable for Chinese micro-blog post, and we make full use of new emoticons which appear frequently in Chinese micro-blog post.

Another significant effort for sentiment classification on Twitter data is by (Barbosa and Feng, 2010). They use polarity predictions from three websites as noisy labels to train a model. They propose the use of syntax features of tweets like retweet, hashtags, link, punctuation and exclamation marks in conjunction with features like prior polarity of words and POS of words. In

order to improve target-dependent Twitter sentiment classification, (Long et al., 2011) incorporate target-dependent features and take the relations between twitters into consideration, such as retweet, reply and the twitters published by the same person. We extend their approach by adding a variety of Chinese dictionaries of sentiment, internet slang and emoticons, and then by using syntactic parser[4] and LIBSVM[5] respectively to achieve the sentiment analysis algorithms based on phrase path and the multiple characteristics for emotional tendency of micro-blog topic. Using micro-blogs' forwarding, commentaries and sharing, we take a future step to optimize the algorithm based on the multiple characteristics.

The problem we address in this paper is to determine emotional orientation for micro-blog topic. So the input of our task is a collection of micro-blogs containing the topic and the output is labels assigned to each of the micro-blogs.

## 3 Data Description

Micro-blog post is a social networking that allows users to post real time messages, and its content is restricted to 140 Chinese characters. Our data set comes from Sina[6] and Tencent[7] micro-blog, and micro-blogs are commonly displayed on the Web as shown in figure 1. "# #" identifies the micro-blog topic, "//" labels user's forwarding relation, "@" specified the user who we speak to, and "V" labeling on the user shows the user's information is identified by the Sina or Tencent.



Fig.1 Micro-blog post example

People usually use sentiment words, internet slang and emoticons to express their opinions and sentiment in micro-blog post. In order to obtain

[1] http://www.tweetfeel.com/

[2] http://twendz.waggeneredstrom.com/

[3] http://teittersentiment.appspot.com/

[4] http://nlp.stanford.edu/downloads/lex-parser.shtml

[5] http://www.csie.ntu.edu.tw/~cjlin/libsvm

[6] http://s.weibo.com/

[7] http://t.qq.com/

emotional orientation on micro-blog topic, we construct some dictionaries described as follows.

## 3.1 The Dictionaries of Sentiment Words and Internet Slang

According to (Du et al., 2009), the sentiment word is one of the best emotional features representations of text, and the rich sentiment words can be conductive to improving emotional tendency identification algorithm. Internet slang that more and more people use in social network is also important factor for emotional orientation. The constructions of them are not only a significant foundation, but also a time-consuming, labor-intensive work.

### 3.1.1 The Dictionary of Sentiment Words

In order to obtain more abundant sentiment words, we regard these sentiment words provided by HowNet[1] and National Taiwan University Sentiment Dictionary (NTUSD)[2] as the foundation, and then use lexical fusion strategy to enrich the dictionary of sentiment words. HowNet is commonsense knowledge base that describes the concepts and reveals the relation between concepts, including the relation between the attributes of concept. Since October 2007, HowNet has released "emotional words collection for sentiment analysis", containing a total of about 17,887 words, of which about 8,942 Chinese words. NTUSD has been summed up by National Taiwan University, including simplified Chinese version and traditional Chinese version, and each version contains 2,812 positive emotion words as well as 8,276 negative emotion words.

(Turney and Littman, 2003; Wang et al., 2011) use lexical fusion strategy to compute the degree of correlation between test word and seed words that have more obvious emotional orientation, and then obtain emotional orientation of test word. We respectively take 20 words with obvious emotional orientation as seed words in this paper, as shown in Tables 1 and 2.

Table 1 Seed words with positive emotion

| 辉煌 | 美妙 | 漂亮 | 俱佳 | 动听 |
|---|---|---|---|---|
| 体面 | 淳美 | 良好 | 出色 | 完美 |
| 美丽 | 精英 | 优秀 | 高手 | 先进 |
| 快乐 | 最佳 | 优质 | 幸福 | 积极 |

Table 2 Seed words with negative emotion

| 罪恶 | 诅咒 | 责备 | 丑陋 | 丑恶 |
|---|---|---|---|---|
| 藐小 | 累赘 | 错误 | 失败 | 麻烦 |
| 不良 | 恶意 | 色情 | 暴力 | 讨厌 |
| 魔鬼 | 野蛮 | 腐败 | 流氓 | 残酷 |

So emotional orientation of the test word is computed as follows:

$$SO(word) = \sum_{pword \in Pset} PMI(word, pword) - \sum_{nword \in Nset} PMI(word, nword)$$

(1)

Where *pword* and *nword* are positive seed word and negative seed word, *Pset* and *Nset* are positive seed words collection and negative seed words collection respectively. *PMI(word₁, word₂)* is described in formula (2), $P(word_1\&word_2)$, $P(word_1)$ and $P(word_2)$ are probabilities of $word_1$ and $word_2$ co-occurring, $word_1$ appearing, and $word_2$ appearing in a micro-blog post respectively. When *SO(word)* is greater than zero, sentiment orientation of word is positive. Otherwise it is negative.

$$PMI(word_1, word_2) = \log(\frac{P(word_1 \& word_2)}{P(word_1)P(word_2)})$$ (2)

### 3.1.2 The Dictionary of Internet Slang

People usually use homophonic words, abbreviated words and network slang to express their sentiment in social network, and (Agarwal et al., 2011) has analysed the sentiment of twitter data. Sometimes new words, produced by important events or news reports, are used to express their opinions. So we construct the dictionary of internet slang to support emotional tendency identification algorithm on micro-blog topic, containing homophonic words, abbreviated words, network slang and many new words.

National Language Resource Monitoring & Research Center (Network Media)[3] has some internet slang, two persons from our lab manually collect more network language through social network, and then integrate these resources together. Finally we achieve this dictionary, containing 861 words with emotional orientation. Table 3 shows part of the dictionary.

---

[1] http://www.keenage.com/html/c_index.html
[2] http://nlg18.csie.ntu.edu.tw:8080/opinion/index.html

[3] http://www.clr.org.cn/

Table 3 Part of the dictionary of internet slang

| Internet slang | Meaning | Polarity |
|---|---|---|
| 达人 | 高人 | Positive |
| 狂顶 | 强烈支持 | Positive |
| 萝莉 | 16岁以下的可爱小女孩 | Positive |
| 灰常桑心 | 非常伤心 | Negative |
| 菜鸟 | 网上低手 | Negative |

## 3.2 The dictionary of emoticons

We construct the dictionary of emoticons by combining emotional symbol library in micro-blog post with other statistical methods. The former is used to select obvious emotion symbols in micro-blog post, such as Sina, Tencent micro-blog et al. The latter choose emoticons used in other social network, containing user-generated emoticons.

Firstly, two laboratory personnel obtain emotional symbol library, and keep the emoticons with the same emotional orientation after their analysis, and then get rid of emotional symbols with ambiguous orientation, the result is described in Table 4.

Table 4 Part of the dictionary of emoticons

| Emoticons | Meaning | Polarity |
|---|---|---|
|  | good | Positive |
|  | 给力 | Positive |
|  | 鄙视 | Negative |
|  | 怒骂 | Negative |

Secondly, in order to enrich the dictionary of emoticons, especially user-generated emoticons in social network, two laboratory personnel collect and analyse emotional orientation, and finally obtain the result shown in Table 5.

Table 5 Part of the dictionary of user-generated emoticons

| Emoticons | Meaning | Polarity |
|---|---|---|
| :o) | 大笑 | Positive |
| :) | 微笑 | Positive |
| :( | 伤心 | Negative |
| T_T | 哭泣 | Negative |

In order to deal with the content conveniently, we pre-process all the micro-blogs and replace all the emoticons with a their sentiment polarity by looking up the dictionary of emoticons.

## 4 Extended Topics

People usually express their sentiment about object by commenting not on the object itself but on some related things of the object. (Yao et al., 2006) expresses the sentiment about automobile by commenting on the attributes or functionalities of the automobile. As shown in the micro-blog post below, user expresses a positive sentiment about Nokia 5800 by expressing a positive sentiment directly about its screen, keyboard et al.

"#诺基亚 5800#屏幕很好，键盘操作也很方便，质量不错哦~亲:)"

It is assumed that user can clearly infer the sentiment about the topic based on those sentiments about the related things. We define those related things as Extended Topics. In order to obtain more micro-blogs' emotional orientation on the topic, we design topic's expansion algorithm, and expand the extended topic set with formula (2). Formula (2) describes the correlation between two words, the greater *PMI* value, the stronger the correlation. So we identify the top 10 nouns and noun phrases which have the strongest association with the topic in micro-blog set, and then take them as the original topic's extended set.

## 5 Sentiment Analysis Based on Phrase Path for Micro-blog Topic

Chinese syntactic analysis has been considered to be an important technique in the process of Chinese information processing. It can be divided into two methods: one is the phrase structure parsing described in (Zhou and Zhao, 2007), namely splits the sentence into phrases, and analyses hierarchical relations among phrases. The other is dependency structure analysis described in (Cheng et al., 2005), namely parse the dependency relations between words. In terms of social network, because the content is brief, and the meaning of the expression is centered relatively, the distance between the sentiment words and evaluation objectives usually is short, and we can use the phrase structure tree to evaluate objectives' emotional orientation.

Take the micro-blog post in section 4 for example, its phrase structure tree is described in figure 2, and its original topic is "诺基亚 5800", we obtain the topic's extended set {屏幕, 键盘, 操作, 质量} using topic expansion algorithm. So

we can confirm the emotional orientation of "诺基亚 5800" by analyzing the sentiment of the element of extended set. At first, starting from evaluation objectives, we find out the shortest phrase path between topic (extended topics) and sentiment word. Phrase path defined in (Zhao et al., 2011) is to link any two phrase nodes in the phrase structure tree. Such as "诺基亚 5800—CD—QP—NP—IP—VP—VP—VA—好", "屏幕—NN—NP—NP—IP—VP—VP—VA—好" et al. Secondly, according to the dictionary of sentiment words, we confirm emotional orientation of the last word in the shortest phrase path. Phrase structure tree is divided into several sub tree by the punctuations, starting from each sub tree, we find out the shortest phrase path, and determine emotional orientation of the topic.
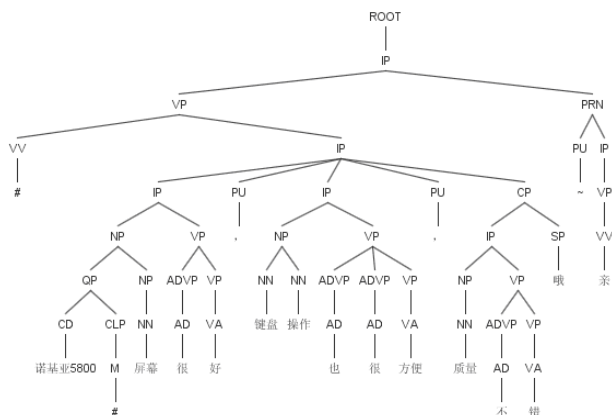


Fig.2 Phrase structure tree of micro-blog example

However, the negative word makes the internal sentiment words tend to shift. Such as "整个店面的装修不是很漂亮", "表演极不自信" et al. "漂亮" and "自信" originally are commendatory terms, but after adding "不是" and "不", the whole sentence semantic changes as a pejorative. In this paper, we adopt matching negative rules method, sentiment words matched by the rules adopt opposite emotional orientation to properly reflect the sentiment. At first, we select negative sentences 6,639 from 20,000 micro-blogs which come from the lab[1], and then accomplish the rule set containing high frequency negative rules 227. Secondly, the rule set is used to match the test sentence, and if sentiment words are the focus of

negation, we take the opposite emotional orientation.

The negative words used in this algorithm are acquired through HowNet. We select the concepts containing negative sememe, such as {neg|否}, {impossible|不会}, {OwnNot|无}, {inferior|不如} et al, and obtain 21 negative words after filtering.

# 6 Multiple Characteristics-based Sentiment Classification for Micro-blog Topics

Based on supervised classification is our another method to determine emotional orientation on micro-blog topic. According to previous related studies (Go et al., 2009; Bermingham and Smeaton, 2010; Agrawal et al., 2003; Pang et al., 2002), NB, ME and SVM are mainly classifiers for text classification, (Go and Barbosa et al.) have ever performed sentiment classification of twitter data, and summarize that SVM is more suitable for short text sentiment classification than other classification models.

We use LIBSVM to achieve sentiment classification based on multiple characteristics for micro-blog topic. LIBSVM is an integrated software for support vector classification, and makes everything automatic—from data scaling to parameter selection. We propose the following procedure:

1) Transform data to the format of an SVM package;
2) Conduct simple scaling on the data;
3) Consider the RBF kernel $K(x, y) = e^{-\gamma \|x-y\|^2}$ ;
4) Use cross-validation to find the best parameter $C$ and $\gamma$ ;
5) Use the best parameter $C$ and $\gamma$ to train the whole training set;
6) Test the whole corpus.

SVM needs categorical features, so we take the features of the Chinese micro-blog post into account. At first, take the data features described in section 3 as some categorical features. According to the polarity of sentiment words, internet slang and emoticons, we compute emotion values and then take these values as some attribute values of SVM. Secondly, take the syntax features of micro-blog content as other categorical features. These syntax features are described as following:

1) Verb-Object structure. Sentiment words are verbs and the topic is their object.
2) Adjective-Center structure, namely a combination of adjective and noun. Sentiment words are adjectives and the topic is their attributive center.
3) Adverb-Center structure, namely an adverb phrase. Sentiment words are adverbs and the topic is their adverbial center.
4) Comparative structure. It is suitable for "A than B + sentiment word" construction. When A is the topic, its emotional orientation is consistent with the polarity of sentiment word; otherwise B is the topic, we adopt the opposite polarity of sentiment word.
5) Emotional orientation shifting. When the emotional orientation shifting is taken place in 1), 2), 3) and 4), we adopt matching negative rules method described in section 5.

According to the polarity of sentiment words, we compute emotion values and then take these values as other attribute values of SVM.

## 7 Sentiment Optimization Based on Relation Features

To some extent, it is limited that only depends on data and syntax features for sentiment analysis of micro-blog topic. In order to improve the accuracy of sentiment analysis, we take relation features between micro-blogs into consideration and then take a future step to optimize sentiment classification based on multiple characteristics.

Micro-blog post is propagated through their forwarding, commentaries, sharing. The forwarding, commentaries and sharing usually means the user agrees with the original user, and they have the same sentiment on the topic. At the same time, micro-blogs published by the same person within a short timeframe[1] should have a consistent sentiment about the same topic. Based on these four kinds of relation features, we can construct a graph using the input micro-blog collection of a given topic. As illustrated in Figure 3, each node in the graph indicates a micro-blog post. The four kinds of edges indicate forwarding (solid line), commentaries (long dash line), sharing relations (dash line) and being published by the same person (round dotted line)

respectively. The isolated nodes have not any relations with other micro-blogs.
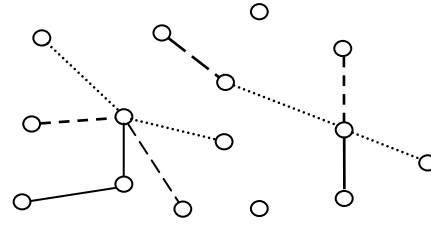


Fig.3 Relationship graph of micro-blogs about a given topic

We consider that the sentiment of a micro-blog post only depends on its content and immediate neighbors, and then compute the sentiment of the current node with the following formula:

$$\Phi_{c,d} = p(\lambda(d)=c \mid \tau(d)) \sum_{\lambda(N(d))} p(\lambda(d)=c \mid \lambda(N(d))) p(\lambda(N(d))) \quad (3)$$

Where $\lambda(d)$ is the sentiment label of node $d$ and value $c \in \{positive, negative, neutral\}$, $\tau(d)$ is the content of node $d$, $N(d)$ is all immediate neighbors of node $d$. The sentiment label of the micro-blog post only depending on its content is represented by $\pi_{c,d} = p(\lambda(d)=c \mid \tau(d))$.

According to relaxation labeling algorithm described in (Angelova and Weikum, 2006), we can simplify the formula (3) into (4), and use formula (4) to iteratively estimate the sentiment for all micro-blogs in the graph.

$$\Phi_{c,d}^{(r)} = \pi_{c,d} \cdot \sum_{\lambda(N(d))} (\prod_{d' \in N(d)} p(\lambda(d)=c \wedge \lambda(d')=c'))^{(r-1)}, \quad r > 1 \quad (4)$$

Where $r$ is iterative superscript. With the shorthand notation $\phi_{c,c'} = p(\lambda(d)=c \wedge \lambda(d')=c')$ we can rewrite this into:

$$\Phi_{c,d}^{(r)} = \pi_{c,d} \cdot \sum_{\lambda(N(d))} (\prod_{d' \in N(d)} \phi_{c,c'})^{(r-1)} \quad (5)$$

The original sentiment label of each node in Fig.3 is computed by the algorithm described in section 6. After the iteration ends, for any micro-blog in Fig.3, the sentiment label that has the maximum $\Phi_{c,d}$ is considered the final label.

# 8 Experiments and Results

## 8.1 Resources and Pre-processing of data

Because there is no annotated micro-blog post corpus publicly available for evaluation of micro-blog topic sentiment classification, we design topic-focused web crawler and respectively crawl 1,000 micro-blogs on three hot topics {iphone5, 袁隆平, 北京爱情故事}. After removing duplicate micro-blogs, we obtain 983 "iphone5", 993 "袁隆平", and 998 "北京爱情故事". Two persons from our lab manually classify each micro-blog post as positive, negative or neutral towards the topics. Among the micro-blogs, 83 of them are neutral-subjective disagreement. In order to manually determine the sentiment for the next step, we assume that these micro-blogs are neutral. 103 of them are positive-negative disagreement, we adopt three people ballot to determine emotional orientation, and finally obtain 1,526 positive, 685 negative and 763 neutral micro-blogs. Take each 500 of them as the training corpus and the other 1,474 as the test set.

## 8.2 The Evaluation of Micro-blog Topic Sentiment Analysis

Sentiment analysis on micro-blog topic is evaluated by Precision, Recall, F-measure and Coverage. The coverage is used to measure integrity and adequacy for the test, and help us understand the test coverage scope.

$$\Pr ecision = \frac{\#system\_correct}{\#system\_proposed} \qquad (6)$$

$$\text{Re} call = \frac{\#system\_correct}{\#person\_correct} \qquad (7)$$

$$F - measure = \frac{2 \times \Pr ecision \times \text{Re} call}{\Pr ecision + \text{Re} call} \qquad (8)$$

$$Coverage = \frac{\#weibo\_topic}{\#weibo\_total} \qquad (9)$$

Where *#system_correct* is the correct result from system, *#system_proposed* is the whole number of micro-blogs from system, *#person_correct* is the number of micro-blogs that has been annotated correctly by people, *#weibo_topic* is the number of micro-blogs containing topic words, *#weibo_total* is the whole number of micro-blogs in the collection.

## 8.3 Results

According to the evaluation in section 8.2, we respectively adopt the algorithms in sections 5, 6 and 7 to determine the emotional orientation for the topics {iphone5, 袁隆平, 北京爱情故事}. As shown in Table 6, we conclude that extending topic is a vital factor, especially it has an obvious effect on sentiment analysis based on phrase path. The mainly reason is that the distance between topic word and sentiment word influences emotional judgment. Extending topic largely shortens the distance and improves the topic coverage, and promotes sentiment analysis performance.

In order to conveniently analyse the results and comparisons, the precision in Tables 7 and 9 is the proportionality of correct micro-blogs identified, containing positive, negative and neutral micro-blogs.

Using sentiment words to determine emotional orientation plays an important role for sentiment analysis. However, as for micro-blog post, the effects of internet slang and emoticons are more significant, the result is described in Table 7. This is because internet slang is simple and flexible to write and expresses rich meanings, emoticons are easy to use and have obvious emotional orientation, they have been used pervasively in social network and welcomed by many Chinese internet users.

Table 7 The influence of internet slang and emoticons for emotional orientation

| Characteristics | Precision（%） |
|---|---|
| Content characteristics | **83.9** |
| - Sentiment words | 81.7 |
| - Internet slang | 75.3 |
| - Emoticons | **71.3** |

However, in order to improve the performance of the evaluation, we take relation features between micro-blogs as an important factor. As seen in Figure 3, there are several micro-blogs which are not connected with any other micro-blogs. For these micro-blogs, our sentiment optimization will have no effect. The following table 8 shows the percentages of the micro-blogs in the collection which have at least one related micro-blog post according to various relation types.

Table 6 The comparisons of algorithms

| Extending topic | Sentiment analysis | Precision（%） | | | Recall（%） | | | F-measure（%） | | | Coverage（%） |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Positive | Negative | Neutral | Positive | Negative | Neutral | Positive | Negative | Neutral | |
| Before extending | Based on phrase path | 80.5 | 53.2 | 23.9 | **43.9** | **31.3** | **73.4** | 56.8 | 39.4 | 36.1 | 68.7 |
| | Multiple characteristics-based | 87.4 | 62.5 | 54.7 | 80.9 | 70.3 | 65.8 | 84.0 | 66.2 | 59.7 | |
| | Based on relation features | **90.8** | 59.1 | 73.0 | 86.7 | 70.3 | 76.0 | 88.7 | 64.2 | 74.5 | |
| After extending | Based on phrase path | 81.8 | 55.0 | 43.9 | **68.5** | **58.9** | **69.6** | 74.6 | 56.9 | 53.8 | 82.6 |
| | Multiple characteristics-based | 92.9 | 69.0 | 69.2 | 83.3 | 78.4 | 90.5 | 87.8 | 73.4 | 78.4 | |
| | Based on relation features | **93.6** | 60.9 | 82.9 | 91.1 | 75.7 | 77.2 | 92.3 | 67.5 | 79.9 | |

Table 8 The proportion of micro-blogs having at least one related micro-blog in the collection

| Relations features | Proportion（%） |
|---|---|
| Forwarding | 39.8 |
| Commentaries | 32.7 |
| Sharing | 22.3 |
| Published by the same person | 25.9 |
| All | **80.3** |

According to Table 8, for 80.3% of the tweets concerning the topics, we can use optimization algorithm to determine their emotional orientation. That means our sentiment optimization based on relation features is potentially useful for most of the micro-blogs. The results reported in Table 9 show that every kind of relation influences on the precision of emotional orientation.

Table 9 Comparison of contributions made by every kind of relations

| Relation features | Precision（%） |
|---|---|
| Forwarding | 67.3 |
| Commentaries | 61.5 |
| Sharing | 78.8 |
| Published by the same person | 70.9 |
| All | **86.7** |
| （Long Jiang et al., 2011） | **83.9** |

As shown in Table 9, compared to 85.6% in (Long Jiang et al., 2011), the precision of our optimization algorithm is increased by more than one percentage points. This mainly attributed to three aspects: firstly, taking internet slang and emoticons as Micro-blog's characteristics is greatly suitable for Micro-blog's sentiment analysis. Secondly, the relations between micro-blogs also play an import role to determine emotional orientation, especially the sharing relation. Thirdly, to some extent the different data set can also influence the precision. We also apply the same data set into (Long Jiang et al., 2011), and the precision is 83.9% which is still lower than our algorithm.

What's more, as a result of micro-blog content published with time characteristic, we can use time characteristic to analyse and forecast micro-blog topic's emotion tendency. We do experiment for the topic "北京爱情故事" published within one week, as shown in Figure 4, the horizontal axis represents time characteristic, and the vertical axis represents the numbers of micro-blogs having emotional orientation.
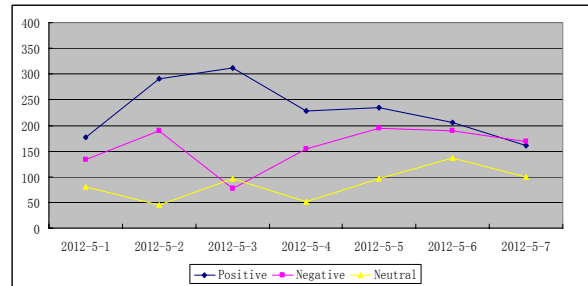


Fig.4 Emotion tendency of the topic "北京爱情故事"

## 9 Conclusions and Future Work

With the emergence of new media, sentiment analysis for new media is becoming more and more important. In this paper, we make full use of multiple characteristics on micro-blog post, and design two kinds of algorithms to achieve micro-blog topic's sentiment analysis. In future the work will focus on the following two aspects:

1) Micro-blog topic's expansion algorithm.
2) The effect produced by the attention relation between users as well as the numbers of fans for micro-blog topic sentiment analysis.

The relations between micro-blog users have equal importance with the multiple characteristics came from micro-blog post itself, they also play a positive effect for micro-blog's sentiment analysis.

## References

Apoorv Agarwal, Xie Boyi, Ilia Vovsha, et al. Sentiment Analysis of Twitter Data [C]. Proceedings of the Workshop on Language in Social Media (LSM 2011). Portland, Oregon, 2011: 30-38

Rakesh Agrawal, Sridhar Rajagopalan, Ramakrishnan Srikant, et al. Mining Newsgroups Using Networks Arising From Social Behavior [C]. Proceedings of the 12th international conference on World Wide Web. Budapest, Hungary: WWW'03, 2003: 529-535

Ralitsa Angelova, Gerhard Weikum. Graph-based Text Classification: Learn from Your Neighbors [C]. Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval. Seattle, Washington, USA: SIGIR'06, 2006: 485-492

Luciano Barbosa and Feng Junlan. Robust Sentiment Detection on Twitter from Biased and Noisy Data. Proceedings of COLING 2010. Beijing, China, 2010: 36-44

Adam Bermingham, Alan Smeaton. Classifying Sentiment in Microblogs: Is Brevity an Advantage? [C]. Proceedings of the 19th ACM international conference on Information and knowledge management. Toronto, Ontario, Canada: CIKM'10, 2010: 1833-1836

Cheng Yuchang, Masayuki ASAHARA, Yuji MATSUMOTOY. Machine Learning-based Dependency Analyzer for Chinese [C]. Proceedings of the International Conference on Chinese Computing 2005. Singapore: COLIPS Publication, 2005: 66-73

Du Weifu, Tan Songbo, Yun Xiaochun, et al. A New Method to Compute Semantic Orientation [J]. Journal of Computer Research and Development, 2009, 46(10): 1713-1720

Alec Go, Richa Bhayani, Huang Lei. Twitter Sentiment Classification using Distant Supervision. CS224N Project Report, Stanford, 2009

Jiang Long, Yu Mo, Zhou Ming, et al. Target-dependent Twitter Sentiment Classification [C]. Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics. Portland, Oregon, 2011: 151-160

Pang B, Lee L, Vaithyanathan S. Thumbs up? Sentiment Classification using Machine Learning Techniques [C]. Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP). Philadelphia, PA: Association for Computational Linguistics, 2002: 79-86

Turney P D, Littman M L. Measuring praise and criticism: inference of semantic orientation from association [J]. ACM Trans on Information Systems, 2003, 21(4): 315-346

Wang Suge, Li Deyu, Wei Yingjie. A Method of Text Sentiment Classification Based on Weighted Rough Membership [J]. Journal of Computer Research and Development, 2011, 48(5): 855-861

Yao Tianfang, Nie Qingyang, Li Jianchao, et al. An opinion mining system for Chinese automobile reviews. In: Cao Youqi, Sun Maosong, eds. Proceedings of the Frontiers of Chinese Information Processing. Beijing: Tsinghua University Press, 2006: 260-281 (in Chinese with English abstract)

Zhou Qiang, Zhao Yingze. Automatic Parsing of Chinese Functional Chunks [J]. Journal of Chinese Information Processing. 2007, 21(5): 18-24

Zhao Yanyan, Qin Bing, Che Wanxiang, et al. Appraisal Expression Recognition Based on Syntactic Path [J]. Journal of Software. 2011, 22(5): 887-898

Zhao Yanyan, Qin Bing, Liu Ting. Sentiment Analysis [J]. Journal of Software. 2010, 21(8): 1834-1848