Connection Science

ISSN: 0954-0091 (Print) 1360-0494 (Online) Journal homepage: https://www.tandfonline.com/loi/ccos20

Sentiment analysis of Chinese microblogging based on sentiment ontology: a case study of '7.23 Wenzhou Train Collision'

Wei Shi, Hongwei Wang & Shaoyi He

Connection Science, 2013
Vol. 25, No. 4, 161–178, http://dx.doi.org/10.1080/09540091.2013.851172

Taylor & Francis
Taylor & Francis Group

Sentiment analysis of Chinese microblogging based on sentiment ontology: a case study of '7.23 Wenzhou Train Collision'

Wei Shi[a,b], Hongwei Wang[a*] and Shaoyi He[c]

[a]*School of Economics and Management, Tongji University, Shanghai 200092, People's Republic of China;* [b]*Business School, Hu Zhou Teachers College, Hu Zhou 313000, People's Republic of China;* [c]*College of Business Administration, California State University, San Marcos, CA, USA*

(*Received 2 April 2013; accepted 18 September 2013*)

Sentiment analysis of microblogging texts can facilitate both organisations' public opinion monitoring and governments' response strategies development. Nevertheless, most of the existing analysis methods are conducted on Twitter, lacking of sentiment analysis of Chinese microblogging (Weibo), and they generally rely on a large number of manually annotated training or machine learning to perform sentiment classification, yielding with difficulties in application. This paper addresses these problems and employs a sentiment ontology model to examine sentiment analysis of Chinese microblogging. We conduct a sentiment analysis of all public microblogging posts about '7.23 Wenzhou Train Collision' broadcasted by Sina microblogging users between 23 July and 1 August 2011. For every day in this time period, we first extract eight dimensions of sentiment (expect, joy, love, surprise, anxiety, sorrow, angry, and hate), and then build fuzzy sentiment ontology based on HowNet and semantic similarity for sentiment analysis; we also establish computing methods of influence and sentiment of microblogging texts; and we finally explore the change of public sentiment after '7.23 Wenzhou Train Collision'. The results show that the established sentiment analysis method has excellent application, and the change of different emotional values can reflect the success or failure of guiding the public opinion by the government.

Keywords: Chinese microblogging; sentiment analysis; fuzzy sentiment ontology; sentiment computing; 7.23 Wenzhou Train Collision

1. Introduction

In recent years, with the fast adaptation of Web 2.0 technologies, microblogging have turned the Web into a vast repository of comments on many topics, generating abundant information that is potentially useful for research in communications, marketing, and social studies. Microblogging, defined by *Oxford Dictionaries Online* (2012) as 'the posting of very short entries or updates on a blog or social networking site, typically via a mobile phone', is an increasingly popular form of communication on the Web. Taking Twitter for example, after signing up with Twitter, a person can post messages via his/her Twitter account and the service distributes them to a limited group of people called contacts and receive messages from them as well. Microblogging posts, commonly known as tweets, are extremely short in comparison to regular blog posts: a post is limited by 140 characters or less. In China, the counterpart of Twitter is Weibo, which is the word for word translation of micro (wei) and blogging (bo) in Chinese. There are four popular

*Corresponding author. Email: hwwang@tongji.edu.cn

microblogging services in China: Sina Weibo, Tencent Weibo, Souhu Weibo, and Netease Weibo. Among them, Sina Weibo is the most popular with more than 300 million registered users in March 2012 (Sina News Center, 2012). Using Chinese celebrities as its public resources for marketing, Sina microblogging services have successfully attracted more and more users to register, the total number of Sina microblogging users had increased by 300% during the year of 2011 (Sina News Center, 2012).

Users of online communities use microblogging to broadcast different types of information. A recent analysis of the Twitter network revealed uses of microblogging for (1) daily chatter, for example, posting what one is currently doing, (2) conversations, that is, directing tweets to specific users in their community of followers, (3) information sharing, for example, posting links to web pages, and (4) news reporting, for example, commentary on news and current affairs (Java, Song, Finin, & Tseng, 2007). Despite the diversity of uses emerging from such a simple communication channel, it has been noted that microblogging normally tend to fall in one of two different content camps: users who microblog about themselves and those who use microblogging primarily to share information (Naaman, Boase, & Lai, 2010). In both cases, microblogging can convey information about the sentiment of their authors. In the former case, sentiment expressions are evident by an explicit 'sharing of subjectivity', for example, 'Hearing this news, I feel terrible'. In other cases, even when someone is not specifically microblogging about their personal emotive status, the message can reflect their mood, for example, 'Signal light problem caused the crash: shocked'. As such, microblogging may be regarded as microscopic instantiations of sentiment.

Chinese are willing to express sentiments on this platform, and microblogging not only has become an important manifestation of the social public opinion, but developments of some important events are strongly influenced by microblogging. Sentiment analysis of microblogging texts becomes increasingly meaningful for government to understand public attitudes and take effective measures to guide the developments. Government departments could make the following arrangements according to the sentiment analysis of Chinese microblogging:

(1) Pay close attention to Spark lines for public sentiment.
(2) On the first day of the incident, respectively, calculate the average value of positive sentiments (expect, joy, and love) and negative sentiments (sorrow, angry, and hate), and compare the two values to determine that whether the incident is positive (the average value of positive sentiments is bigger than the average value of negative sentiments) or negative (the average value of negative sentiments is bigger than the average value of positive sentiments).
(3) Government departments could take different interventions according to the polarity of the incident.
(4) In the development process of the incident, compare the changes of different sentiment values over time and take different strategies depending on the changes (e.g. when value of negative sentiments decreases, indicating that the intervention measures are effective, so the government could continue to strengthen this measure).
(5) Determine the normal value of sentiment, when the value of different sentiments stays on a certain level, indicating that the incident calms down, the government could reduce the intervention.

In this article, we explore how public sentiment patterns can be modelled and evaluated with a sentiment analysis of microblogging posts published in Sina Weibo between 23 July and 1 August 2011 about '7.23 Wenzhou Train Collision'. On 23 July 2011, two high-speed trains collided on a viaduct in the suburbs of Wenzhou, Zhejiang Province, China. The two trains derailed each other, having 40 people killed and at least 192 injured. The derailed cars were buried hastily and the rescue operations were concluded quickly at the orders of government officials. Such actions had caused very strong criticism from Chinese media and online communities, including the microblogging community. Tens of thousands microblogging posts were published on the four major microblogging platforms: Sina Weibo, Tencent Weibo, Souhu Weibo and Netease Weibo.

In our study, first, we will collect all of the microblogging posts about '7.23 Wenzhou Train Collision' published on Sina Weibo services over the 10-day period, taking each microblogging post as a short Web-based text; second, we will measure the public's sentiment state with our sentiment space model which was based on the eight sentiment classes, namely expect, joy, love, surprise, anxiety, sorrow, angry, and hate; third, we will conduct an analysis of microblogging text based on fuzzy sentiment ontology. With our analysis, we attempt to assess whether a quantifiable relationship between overall public sentiment and social events can be identified.

2. Literature review

An increasing number of empirical analyses of sentiment and mood are based on textual collections of public user-generated data on the Web.

2.1. *Sentiment analysis*

Sentiment analysis, also known as opinion mining, is to find consumers' attitudes and opinions about goods through automatic analysis of texts of commodity reviews. Stone, Dunphy, and Smith (1966) introduced and defined a model of content analysis, presented a classification system for many areas such as political science, personality, social psychology, product image, and so on. Their research could be foundational work for sentiment analysis. Sentiment analysis has different versions: broad and narrow. Broadly speaking, sentiment analysis includes analysis of a variety of emotions in the text; narrowly speaking, sentiment analysis refers to views of judging things or themes by users, for example, 'agree' or 'against' (Pang, Lee, & Vaithyanathan, 2002). Current opinion mining technology is divided into two classes: machine learning and semantic orientation (Chaovalit & Zhou, 2005a). Pang et al. (2002) proposed machine learning method to mine sentiment tendencies, and the accuracy of their method achieved 87.5%. Some followed scholars extended their study and also got good results in experiments. Turney (2002) proposed Semantic Orientation from Pointwise Mutual Information semantic classification method based on the emotional phrases, and its reliability has got a preliminary validation in the study of sentiment classification. Dave employed the method to take sentiment analysis of customer reviews in the Amazon and C-Net and other online stores; it had once again proven the performance of the method (Dave, Lawrence, & Pennock, 2003a). Chaovalit and Zhou (2005b) used the data of film reviews to compare the sentiment classification method based on semantic orientation with the sentiment classification method based on machine learning, and he found that the results of semantic methods are similar to machine learning methods. Wang, Yin, Yao, & Liu (2012) studied on sentiment classification of Chinese online reviews, and they adopted four statistical feature selection methods to select features and employed a support vector machine classifier to predict the sentiment polarity of online reviews; they also proposed an ontology-based linguistic model to identify the basic appraisal expression in Chinese product reviews – 'feature–opinion pair', and their research contributed to the related research in opinion mining (Wang et al., 2012; Yin, Wang, & Guo, 2013).

In addition to academic research, some scholars began to study the relevant applications of sentiment analysis. Eguchi and Lavrenko (2006) used a model of sentiment retrieval in his study, enabled users to query according to certain emotion words, themes, and attitudes. Dave, Lawrence, & Pennock (2003a) distinguished the sentiments for the product reviews and developed the world's first sentiment analysis tool Review Seer. Mishne and Glance (2006) predicted movie sales from blogger sentiment. Coppola UK software company developed the 'emotional' software, which could judge whether the attitude of articles or reviews published on medias towards the government's policy was positive or negative, and obtained the conclusion in a very short time (Global Times, 2005).

2.2. *Sentiment classes*

Human sentiments (attitude) are complex and constantly changing. Osgood and Tannenbaum (1955) described a general theory of attitude change. Predicted changes in attitude towards both source and concept were based upon the combined operation of a principle of congruity, a principle of susceptibility as a function of polarisation, and a principle of resistance due to incredulity for incongruous messages. They indicated that there were many variables which contribute to attitude change. So far, there is no recognised classification of human sentiments, and there are 6, 8, 10, or even two dozen classifications of sentiment classes, as the study on the sentiment classes is progressing and developing. Existing classification methods are listed in Table 1.

Among them, Changqin and Fuji (2010) studied on emotional expression in Chinese blog; they selected eight emotion classes (expect, joy, love, surprise, anxiety, sorrow, angry, and hate) for their emotional expression analysis and achieved good results. Chinese microblogging are similar to blogs in emotional expression. In order to reduce confusion on selecting of sentiment class and include most common sentiment in microblogging, we selected the same eight sentiment classes (expect, joy, love, surprise, anxiety, sorrow, angry, and hate) for the construction of fuzzy sentiment ontology and sentiment analysis of Chinese microblogging.

2.3. *Construction of sentiment vocabulary*

In early stage, Sparck Jones (1965) considered the problem of constructing a thesaurus, and this involved a method for defining the meanings or uses of words, and a procedure for classifying them; it was suggested that word uses may be defined in terms of their 'semantic relations' with other words and that the classification may be based on these relations. Recently some scholars tried to express multiple sentiment orientations of words through sentiment vocabulary ontology database. Yan, Bracewell, Ren, and Kuroiwa (2008) constructed Chinese sentiment ontology based on the HowNet, and finally, the ontology includes 5500 verbs which have 113 different emotional classes. Subasic and Huettner (2001) considered that the word not only has one sentiment tendency, but it may be related to one or more sentiment classes. The ultimately sentiment tendency depended on the centrality and intensity of the atoms sentiment words, and they summed up 83 atoms sentiment classes (Subasic & Huettner, 2001). Liu et al. proposed to use open mind common-sense database to give emotional value to the selected language features. They summarised six basic classes (happy, sad, angry, fear, disgust, and surprise) and finally determined the polarity of the text by analysing the emotional words (Liu, Hu, & Cheng, 2005). Pan and Lin got the sentiment words through establishing sentiment vocabulary ontology, and they divided sentiment into 20 classes. The sentiment vocabulary ontology was described through a triple, Lexicon $= (B, R, E)$, where E was the sentiment class and intensity of vocabulary (Pan & Lin, 2008). Narisa Zha used basic fuzzy evaluation set to do the conceptual model description of class, intensity, and polarity of sentiment vocabulary ontology. They employed the basic fuzzy evaluation set to manually annotate emotional and evaluation words in corpus, and determined the sentiment class and intensity of the sentiment words by the method of several people annotated and analysed together (Na, Liu, & Li, 2010). Application of the ontology opens a new direction for analysis of sentiment orientation for words (Zhang, Li, Ren, Li, & Kuroiwa, 2005).

Table 1. Summary of sentiment classes.

Author/year	Research field	Sentiment class	Specific class	Languages
Ekman (1993)	Facial expressions	Six classes	Enjoyment, sadness, fear, disgust, anger, surprise	English
Xu and Tao (2003)	Expression system in Chinese	Seven classes	好（爱，敬） good (love, respect),恶 (evil), 喜（乐）joy (happy), 怒 (anger), 哀 (sadness), 惧 (fear), 欲 (desire)	Chinese
Xu and Tao (2003)	Based on psychological feelings and on expression	Two classes	*Psychological feelings*:喜 (joy), 乐 (happy), 爱 (love), 愁 (depression), 闷 (nausea), 悲 (grief), 慌 (panic), 敬 (respect), 激动 (excitement), 羞 (shame), 疚 (remorse), 烦 (tired), 急 (anxious), 傲 (proud), 吃惊 (surprised), 怒 (angry), 失望 (disappointed), 安心 (peace of mind), 恨 (hate), 嫉 (envy), 蔑视 (contempt), 悔 (regret), 委屈 (grievances), 谅 (understanding), 信 (faith), 疑 (doubt). *Expression*: 态度词 (attitude words), 品性词 (character words), 声音词 (sound words)	Chinese
Zhang et al. (2005)	Emotion vocabulary	12 classes	高兴 (happy), 悲哀 (sad), 恐惧 (fear), 厌恶 (disgust), 愤怒 (anger), 惊奇 (surprise), 喜爱 (love), 期待 (expect), 焦虑 (anxiety), 内疚 (guilt), 赞扬 (praise), 羞 (shame)	Chinese
Lin (2006)	Da Xubeng <says the text> the words of sentiment expression	18 classes	安静 (quiet), 喜悦 (joy), 恨怒 (anger), 悲痛 (grief), 哀怜 (pity), 忧愁 (sorrow), 忿急 (anxious), 烦闷 (depressed), 恐惧 (fear), 惊骇 (horror), 恭敬 (respect), 抚爱 (caress), 憎恶 (hate), 贪欲 (greed), 嫉妒 (jealousy), 骄慢 (arrogance), 惭愧 (shame), 耻辱 (humiliation)	Chinese
Xu and Ling (2007)	'Seven emotion' considering the Chinese tradition	Seven main classes and 20 categories	乐 (happy), 好 (good), 怒 (angry), 哀 (sad), 惧 (fear), 恶 (evil), 惊 (scared) 快乐 (happy), 安心 (peaceful), 尊敬 (respect), 赞扬 (praise), 相信 (trust), 喜爱 (love), 愤怒 (angry), 悲伤 (sad), 失望 (disappoint), 思 (thinking), 慌 (panic), 恐惧 (fear), 羞 (shame), 烦闷 (boring), 憎恶 (hate), 贬责 (deprecating), 妒忌 (jealousy), 骄慢 (arrogance), 怀疑 (doubt), 惊奇 (surprise)	Chinese
Changqin and Fuji (2010)	Emotional expression in blog	Eight main classes	Expect, joy, love, surprise, anxiety, sorrow, angry, hate	English

2.4. *Sentiment analysis of microblogging*

There have been more and more scholars who are very interested in studying microblogging as a subject for sentiment analysis and opinion mining (Barbosa & Feng, 2012; Bollen, Gonçalves, Ruan, & Mao, 2011; Bollen, Pepe, & Mao, 2012; Cheong & Lee, 2011; Diakopoulos & Shamma, 2010; Go, Bhayani, & Huang, 2009; Gruzd, Doiron, & Mai, 2011; Jiang, Yu, Zhou, Liu, & Zhao 2011; Kouloumpis, Wilson, & Moore, 2011; Pfeiffer & Tourte, 2012; Thelwall, Buckley, & Paltoglou, 2011; Zhang, Ghosh et al., 2011). Various researchers have conducted sentiment analyses using Twitter as a corpus from different perspectives. Kouloumpis et al. (2011) looked at the linguistic features of Twitter posts to evaluate the usefulness of existing lexical resources and their features, and found out that part-of-speech (POS) features might not be useful for analysing the sentiment revealed in microblogging platform such as Twitter. In their attempt to classify sentiment in microblogs, Berminghan and Smeaton (2012) conducted an experiment on a collection of over 60 million Twitter posts and found out that it was much easier to classify the sentiment in microblogging despite the sparsity of information provided by the short messages posted on Twitter. Davies and Ghahramani (2011) proposed a language-independent probabilistic model for classifying happy vs. sad sentiment in short, social network statuses based on data from Twitter, showing that this model outperformed Naive Bayes models by more than 10% when there are additional unlabelled training examples.

In China, since its appearance in late 2009 as a counterpart of Twitter in the name of weibo (the Chinese translation of microblogging), microblogging had quickly gained more and more popularity as a form of social media favourably used by Chinese online communities. According to a media research report done in March 2011, microblogging had become the third favourite online source of information, after news portals and online forums (Cheng, 2011). Along with such a rapid development of microblogging as a nation-wide phenomenon in China, there also has been a rapid development of microblogging research with substantial achievements in different research areas, such as communication studies, marketing research, information retrieval, political science, and user studies (Collett, 2012; Ding, 2011; Huang, Chan, & Hyder, 2010; Liu, Björkstén, & Lagerdahl, 2012; Qu, Huang, Zhang, & Zhang, 2011; Yu, Asur, & Huberman, 2011a; 2011b; Zhang, Sun, et al., 2011; Zheng, Ren, Liu, & Xu, 2011).

Sentiment analysis techniques rooted in machine learning yield accurate classification results when sufficiently large data are available for testing and training. Minute texts such as microblogs may make particular challenges for this approach, since the semantic approach, which does not require training and testing, may enable sentiment analysis for very small text data. In this study, we will analyse the public sentiment in microblogging by the fuzzy sentiment ontology combined with the semantic factor.

3. Research methodology

Our methodology is depicted in Figure 1 and consists of the following steps:

(1) Definition of data and sentiment assessment instrument.
(2) Construction of fuzzy sentiment ontology.
(3) Data cleaning, parsing, and normalisation.
(4) Calculation of text influence.
(5) Calculation of sentiment value in the microblogging.
(6) Comparison of produced sentiment time series to the dynamic of '7.23 Wenzhou Train Collision'.

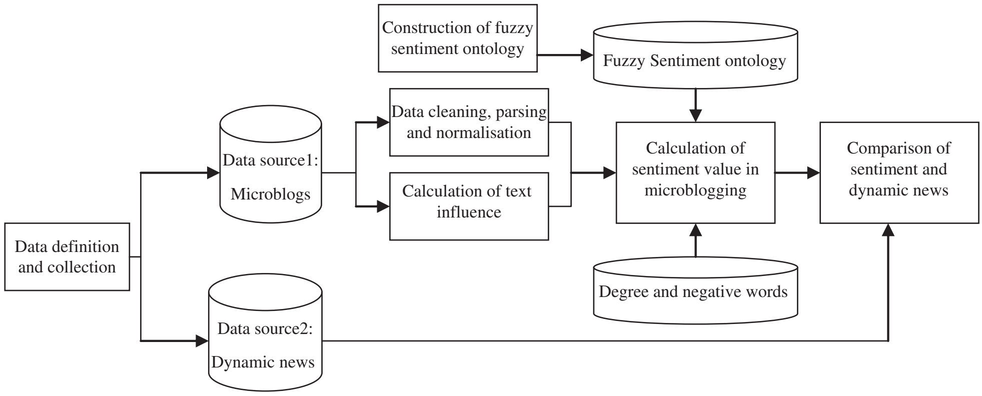

Figure 1. The system overview graph of our methodology.

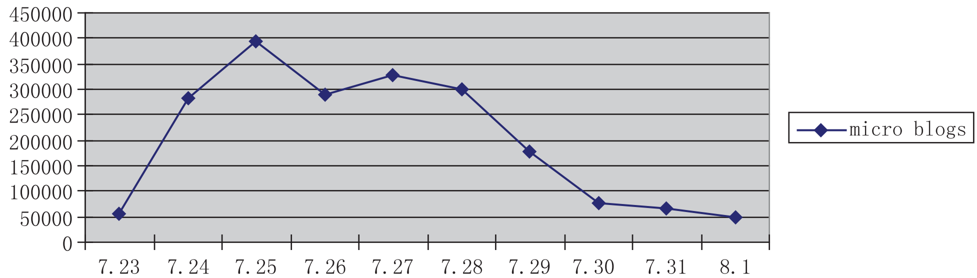

Figure 2. Daily microblogs distribution: 23 July to 1 August 2011.

3.1. *Data collection*

Our analysis is based on two data sources. The first one is the manually compiled news of ‘7.23 Wenzhou Train Collision’ that took place during the time period of 23 July to 1 August 2011. All of the dynamic news information comes from the important Chinese Web portal Sina.com (http://www.sina.com.cn/).

Second, we collected all public microblogs on ‘7.23 Wenzhou Train Collision’ posted by Sina Weibo users from 23 July to 1 August 2011. The resulting data set consists of 2,024,792 microblogs temporally distributed as shown in Figure 2.

From Figure 2 we could find that the number of microblogging posts on the ‘7.23 Wenzhou Train Collision’ showed a sharp increase in the trend from the evening of 23 July. On 25 July the number of posts on the ‘7.23 Wenzhou Train Collision’ achieved 395,808 in Sina Weibo. The number of post has dropped slightly to 288,344 on 26 July. However, the number of posts has increased by 327,712 on 27 July. Then, the daily number of microblogging posts gradually decreased. On the 10th day (8.1) the number dropped down to 50,540. These changes are complying with the law of development of hot events in the network.

3.2. *Construction of fuzzy sentiment ontology*

The sentiment ontology aims to identify recognised terminologies of sentiment in users’ reviews, obtain clear definitions of these terminologies and relations among them, achieve a common understanding of sentiment, and realise the sentiment analysis on online reviews at semantic level.

This paper takes the words set for sentiment analysis of HowNet as a source of sentiment vocabulary. The fuzzy sentiment ontology which is constructed by the paper could be the basis of sentiment analysis of online reviews in the future.

The fuzzy sentiment ontology can be described by a 3-triples, as

$$\mathrm{FSO} = (B, R, E).$$

Here, B is the basic information of vocabulary, including number, term, the vocabulary in English, POS, entry, and version information; R represents the synonymous relationship between terms; and E is the emotion class and intensity of vocabulary:

B (ID ; term ; term in English; POS; entry; version)
R(synonymous vocabulary)
E(sentiment class ; membership)

Such as FSO = ((8; 仰慕; admiration; V; Wei Shi; HowNet, 2007)

(敬慕；爱慕；景仰)

(喜爱；0.44)).

(1) *ID* (Identity): The unique number of sentiment vocabulary for indexing.
(2) *Term*: The word that expresses some sentiment. Sentiment vocabulary is obtained from HowNet (2007) as sources of sentiment ontology. They will be continually updated as massive new sentiment words have been emerging on Internet.
(3) *Term in English*: English words which is translated from corresponding Chinese vocabulary.
(4) *POS*: The POS of a word varies with contexts, thus the word with different POS would express different emotion class and intensity. Here, we use Institute of Computing Technology, Chinese Lexical Analysis System, a Chinese lexical analysis system, for POS tagging.
(5) *Entry*: The person or group who completes the entry is recorded for facilitating statistics and verification.
(6) *Version*: The version of sentiment ontology.
(7) *Synonymous vocabulary*: The words that have the same sentiment class and sentiment intensity with the term.
(8) *Sentiment class*: The classes of emotion words are divided into eight classes, as follows: expect, joy, love, surprise, anxiety, sorrow, angry, and hate. There are positive and negative emotions, which we call emotional polarity, expect, joy, and love are positive emotions in eight emotion classes, and sorrow, angry, and hate are negative emotions, while surprise and anxiety may express both negative and positive in different contexts. We will extract the sentiment vocabulary to compare with the eight basic emotion classes, respectively, and the sentiment class which has the largest membership will be the class of the sentiment vocabulary.
(9) *Membership (intensity)*: The degree of membership, ranging between 0 and 1, denotes the sentiment intensity of evaluation words or emotion words.

Human sentiments are fuzzy and complex. The reviews in microblogging are users' emotional expression for something or someone. In this paper, sentiment of a microblogging text is expressed by a vector:

$$\overrightarrow{d} = \langle e_1, e_2, \ldots, e_n \rangle. \tag{1}$$

Here, $e_i (1 \leq i \leq n)$ is a basic sentiment class contained in document d. The values of e_i range from 0 to 1 (discrete), indicating the intensities of the basic sentiment classes.

Table 2. The different membership of 'happy' for different sentiment classes.

Sentiment classes	期 待 (expect)	高 兴 (joy)	喜 爱 (love)	惊 讶 (surprise)	焦 虑 (anxiety)	悲 伤 (sorrow)	生 气 (angry)	讨 厌 (hate)
开 心 (happy)	0.24	1.00	0.21	0.29	0.04	0.29	0.29	0.21

HowNet is an online common-sense knowledge base unveiling inter-conceptual relations and inter-attribute relations of concepts as connoting in lexicons of the Chinese and their English equivalents. For Chinese vocabulary, the description in HowNet is based on the basic concept – 'sememe'. A sememe refers to the smallest basic semantic unit that cannot be reduced further in Chinese. As the meaning of 'word' in Chinese is very complex, a word often expresses different semantics in different contexts. Thus, the Chinese word could be understood that it is the collection of several senses in HowNet. In semantic dictionary of HowNet, each record is the composition of one sense and its description of one word that a record corresponds with a sense of a word. For example, '舒服' (comfortable) has three senses: V Be Well | 健壮; ADJ a Value | 属性值, circumstances | 境况, peaceful | 和平; V satisfied | 满意. In recent years, many scholars use the structure and semantics resources provided by HowNet to study on semantics in Chinese.

This paper takes 836 positive emotion words and 1254 negative emotion words provided by sentiment vocabulary of HowNet 2007 version as the corpus source of fuzzy sentiment words ontology (http://www.keenage.com). An emotion word has different membership for different sentiment classes, as shown in Table 2.

Here since selecting the basic sentiment class which has the largest membership as the sentiment class of the emotion vocabulary, the largest value of membership is the intensity of the emotion vocabulary, and here the intensity of "开心 (happy)' is 1.00. If there are some same memberships, the most appropriate emotion class is selected manually. This sentiment class of "开心 (happy)' selects "高兴 (joy)'.

In prior studies, the process of determining membership is mainly dependent on manual judgement from linguists. Thus, the process would inevitably be affected by some subjective factors, such as personal experiences, educational background, and so on. Some studies try to minimum the impact caused by personal factors by asking for more participants from different fields, but the result fails to be satisfied. As the emotional words themselves are similar in the semantics and structure, we can obtain the membership by calculation of semantic similarity of sentiment words to each basic sentiment class. The value of semantic similarity in HowNet ranges between 0 and 1.

In the process of the semantic similarity calculation, the vocabulary semantics can be represented by sememe in HowNet. All sememes constitute a tree hierarchy according to their hyponymy, and semantic similarity of two sememes can be got by semantic distance similarity approach (Equation (3)). The vocabulary semantics is divided into four parts: 'the first basic sememe', 'other basic sememe', 'relationship sememe', and 'symbolic sememe'; similarity calculation between vocabulary semantics s_1 and s_2 is as follows:

$$\mathrm{sim}(s_1, s_2) = \sum_{i=1}^{4} \beta_i \prod_{j=1}^{i} \mathrm{sim}_j(s_1, s_2), \tag{2}$$

$$\mathrm{sim}_j(s_1, s_2) = \frac{\alpha}{d + \alpha}, \tag{3}$$

where β_i ($1 \leq i \leq 4$) are the adjustable parameters, $\beta_1 + \beta_2 + \beta_3 + \beta_4 = 1$, $\beta_1 \geq \beta_2 \geq \beta_3 \geq \beta_4$, $\mathrm{sim}_j(s_1, s_2)$ is one similarity of the four sememes, d is the distance of two sememes in tree hierarchy, and α is a adjustable parameter.

Figure 3. The dialog box of lexical similarity calculation.

The calculation program of vocabulary semantic similarity, which is written according to the principle in Liu and Li (2002), realises the calculation of semantic similarity between the sememes automatically. The calculation of words similarity can be achieved on the basis of the calculation of sememes similarity, since words similarity is defined as the maximum similarity in all meanings similarity of two words. By entering two words and then selecting the exact meanings, respectively, the numerical value of similarity would be obtained in the box of results. The similarity value is the membership value of an emotional word. Specific dialogue interface is shown in Figure 3. For example, given an emotion word '称心 (bed of roses)' and an emotion class '高兴 (joy)', the output will be 0.44 if we select the corresponding meanings '{satisfied | 满意}' and '{joyful | 喜悦}'. It means the semantic similarity between the two words under the corresponding meaning is 0.44, and the value is also called the degree of membership of '称心 (bed of roses)' to '高兴 (joy)'. Here α, β_1, β_2, β_3, β_4, γ, and δ are adjustable parameters for making the degree of membership more reasonable. α is an adjustable constant for the first basic sememes, γ is a similarity constant of the specific words and sememes, and δ is a similarity constant of any non-empty value and null value.

For example, a fuzzy sentiment words ontology
FSO = ((18; 开心; happy; adj; Wei Shi; HowNet, 2007 版情感分析用词语集)
(快乐); 愉快)
(高兴; 1.00)).

The final fuzzy sentiment words ontology include 2090 words, which belong to the eight basic sentiment classes as shown in Table 3.

Table 3. The number of each kind of sentiment words.

Sentiment class	Expect	Joy	Love	Surprise	Anxiety	Sorrow	Angry	Hate
Number of words	170	395	339	65	271	220	201	429

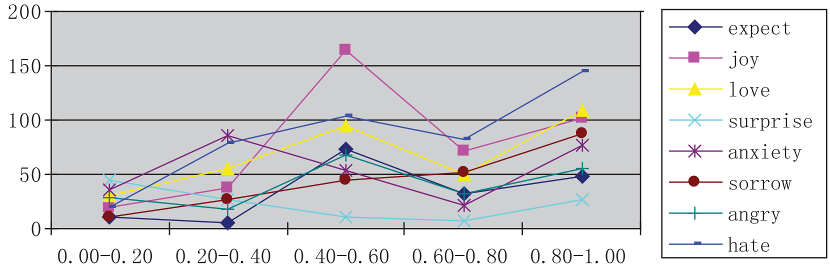

Figure 4. The intensity distribution of sentiment words.

Table 4. Accuracy rate of sentiment class determination.

Method	PMI method	Our method
Accuracy	82%	86%

The intensity distribution of sentiment words is shown in Figure 4; overall, the number of sentiment words whose intensity are in 0.40–0.60 and 0.80–1.00 is more than other sections.

The paper takes PMI (pointwise mutual information) in the literature (Hong, Hongfei, & Yu, 2008) to compare with our method. The literature (Hong et al., 2008) used PMI to get the emotion class and intensity of emotion words in the construction of sentiment ontology. The comparison results of PMI and our method is given in Table 4.

The results show that our method has higher accuracy in emotion class determination; moreover, our method has higher efficiency in practice.

3.3. *Text cleaning and parsing*

Since the 2,024,792 microblog posts on the '7.23 Wenzhou Train Collision' collected from Sina Weibo did not all directly express emotions, this study employed semantics-based approach to study microblogging sentiment, and only retained microblogging posts which contained 2090 words in fuzzy sentiment ontology. Consequently, this procedure leads to a reduction of posts from 2,024,792 to 1,039,273, containing mostly expressions of individual sentiment states. These posts are normalised and parsed before processing as follows:

(1) Separation of individual terms on white-space boundaries.
(2) Removal of all non-alphanumeric characters from terms, for example, commas, dashes, etc.
(3) Removal of 1208 standard stop words, including highly common verb forms.
(4) In order to avoid spam messages and other information-oriented posts, filter out posts that match the regular expressions 'http:' or 'www.' and the user's name (marked with the @ symbol).
(5) Remove the 'Reply', 'forward microblogging', and other words and forwarded content (just forward did not increase any comment posts).
(6) After clean-up, the microblogging is divided into a single sentence, then basic POS tagging and sentiment words annotation.

3.4. *Calculation of text influence*

The text influence of a microblogging is related to posters (Hou & Zhao, 2011), and they consider this from several factors: (1) whether the user is the one who has been authenticated by the microblogging platform, if the user is a social celebrity, he has a strong influence; (2) the number of followers for the user, more followers indicate that he is more influential; and (3) the number of the user's friends, if the number of the user's friends is excessive, then he is just the recipient of information, the text influence is very small. Considering these factors, they get the calculation methods of influence for text y, as follows:

$$w_y = x\left(\frac{f_{y,1}}{f_{y,2}}\right) \times \min\left\{\frac{f_{y,1}}{\delta}, \frac{f_{y,1}}{f_{y,2}}\right\} \times \nu, \tag{4}$$

where the term $f_{y,1}$ is the number of followers of the user who publishes text y and $f_{y,2}$ is the number of friends of the user who publishes text y. The function $x(l)$ is the expansion scale coefficient of the influence; according to the characteristics of the microblogging platform it is defined as follows: when $l \geq 10, x(l) = 2$; when $1 < l \leq 10, x(l) = 1$; when $l \leq 1, x(l) = 0$. δ, ν are the adjustable constants, where δ is a set depending on the microblogging platform; ν denotes whether the user is authenticated by the microblogging platform, if he has been authenticated, $\nu > 1$, otherwise, $\nu = 1$.

3.5. *Calculation of sentiment value in m icroblogging*

In the calculation of sentiment value, microblogging texts are annotated according to the established fuzzy sentiment ontology database, for example:

'愤怒的群众，动车事故引发中国网民集体愤怒。' (Raging crowds, 7.23 Wenzhou Train Collisions caused the collective rage of Internet users in China.)

There is one sentiment class in this microblogging text annotating the following:

'愤怒(raging)' V 生气(anger) 1.00

The degree words are related to the intensity of emotion. They often appear together with sentiment words and thus change the intensity of the sentiment words. In order to better analyse the emotional intensity of microblogging texts, we set a detection window in the context of sentiment words, and window size is 5 in our experiment. If degree words appear within the detection window, correspondingly the emotional intensity of sentiment words will increase from 1.5 to 0.8 according to different grades of degree words. This paper extracts 60 degree words from the HowNet and divides them into seven classes, the setting is listed in Table 5 (as the complexity of Chinese, some Chinese degree words have the same English translation, while their Chinese meanings are different).

Negative words have been frequently used in Chinese. Analysis on negative words will be quite helpful for accurate sentiment analysis. Twenty-two negative words are manually extracted from HowNet in this paper. We also set a detection window in the context of sentiment words. If the negative word appears in the detection window, the emotional value of the phrase is negated. The size of window is 5 in our experiment.

When there are a number of sentiment vocabularies which belong to the same sentiment class in a sentence, the intensity of sentiment class is the average of the intensity of these sentiment vocabularies.

According to the rules above, the sentiment value in microblogging y is calculated as follows:

$$e_i = w_y \times \frac{\sum_{m=1}^{m} (-1)^n \times \text{value}_{\text{deg}} \times \text{Sensibility}(k_m)}{m} \quad \text{if } k_m \in e_i, \tag{5}$$

where $1 \leq i \leq 8$.

Here, $\text{Sensibility}(k_m)$ indicates the original value of sentiment vocabularies k_m (membership of sentiment class) $0 \leq \text{Sensibility}(k_m) \leq 1$; n represents the occurrence number of negative words

Table 5. Assignment value of degree words.

Assignment value	Degree words
1.5	最(bottom)、最为(most)、极(mighty)、极为(very)、极其(spanking)、极度（to the utmost）、极端(exceeding)
1.4	太（so much）、绝(absolutely)、至为(to the)、顶(top)、(over)、过于(excessively)、过分(overmuch)、分(exceptionally)、万分(extremely)、何等(how)
1.3	很(quite)、挺(rather)、怪(odd)、老(always)、非常(greatly)、特别(special)、相当(quite)、十分(great)、甚(very)、甚为(very)、异常(remarkably)、深为(deeply)、蛮(pretty)、满(completely)、够(really)、多(much)、多么(so)、殊(outstanding)、何其(how)、尤其(especially)、无比(unequaled)、尤(particularly)、超(super)
1.2	不甚(fully)、不胜(extremely)、好(fine)、好不(no better)、颇(considerably)、颇为(quite)、大(big)、大为(much)
1.1	稍稍(slightly)、稍微(a little)、稍许(slightly)、略(slightly)、略为(slightly)、多少(how much)
0.9	较(relatively)、比较(comparatively)、较为(more)、还(also)
0.8	有点(a little)、有些(some)

in the detection window; m is the number of words which belong to the same sentiment class in a microblog; value$_{\text{deg}}$ represents the intensity value of degree words, previously described; w_y is the influence factor of microblogging text y which has been introduced in Section 3.4; and finally, each microblogging text could been represented by $\overrightarrow{y} = \langle e_1, e_2, \ldots, e_i, \ldots, e_8 \rangle$.

We produce $e_{i,d}$ for the value of sentiment class i of microblogs submitted on a particular date d, denoted $T_d \subset T$ by simply averaging the sentiment value of the microblogs submitted that day, that is,

$$e_{i,d} = \frac{\sum_{\forall t \in T_d} e_i}{\| T_d \|}. \tag{6}$$

Here, $\| T_d \|$ is the valid number of microblogs on date d. e_d is the sentiment vector expressed by microblogging on date d (the set of $e_{i,d}$); the time series of aggregated, daily mood vectors e_d for a particular period of time$[j, k]$, denoted $\theta_{e_d}[j, k]$, is then defined as

$$\theta_{e_d}[j, k] = [e_j, e_{j+1}, \ldots, e_d, \ldots, e_k].$$

Since the daily number of microblogs is different (as shown in Figure 1), this leads to systemic changes in the variance of $\theta_{e_d}[j, k]$ over time. In particular, the variance is larger in the early or late days of the incident, when microblogs are relatively scarce. As the number of microblogs is more in the medium term of the incident, the variance of the time series is smaller. This effect makes it problematic to compare changes in the sentiment vectors of $\theta_{e_d}[j, k]$ over time. We adopt a variable mean so that we can make comparisons of general sentiment levels between different periods of time while maintaining normalised variance; we define the variance-normalised sentiment vector as follows:

$$\bar{e}_d = \frac{e_d}{\sigma(\theta[j, k])}, \tag{7}$$

where $\sigma(\theta[j,k])$ represents the standard deviation of sentiment vector of microblogging within the local $[j, k]$, and consequently, we obtain the following variance-normalised time series:

$$\bar{\theta}_{e_d}[j,k] = [\bar{e}_j, \bar{e}_{j+1}, \ldots, \bar{e}_d, \ldots, \bar{e}_k]. \tag{8}$$

4. Experiments and data analysis

Our case study is to analyse the change of public sentiment expression on the Sina Weibo platform in the 10-day period after 'Wenzhou 7.23 Wenzhou Train Collision' at 20:34 on 23 July 2011. Sentiment value of 1,039,273 microblogs are computed according to the formulas (5)–(7), the change of sentiment values are shown in Table 6, the curves are shown in Figure 5, the abscissa represents the date (day), and the vertical axis indicates the sentiment value of each sentiment class.

Referring to the dynamic news on '7.23 Wenzhou Train Collision' and Figure 5, we could find the following:

The first day of '7.23 Wenzhou Train Collision' (7.23): 'expect' and 'anxiety' occupied on peak in a variety of sentiments, the microblogs are full of the sentiment of concerns about passengers and expectations for a miracle when people just heard the sad news and hoped that more people were rescued; the value of 'Sad' arrived 0.49, many people expressed sad feeling about fragile life and some mainstream medias did not timely report the incident in the microblogging; the value of 'joy' and 'love' was in a very low level, this was understandable.

The second day of '7.23 Wenzhou Train Collision' (7.24): the value of 'anxiety' was the highest, with the rescue work, the public began to worry about dead and injured passengers and be anxious about the relief work; the value of 'sorrow', 'surprise', 'angry', and 'hate' increased significantly, refer to 7.24 Event Dynamic, the public expressed shock on the rescue work (buried the front of

Table 6. The change of sentiment value.

	7.23	7.24	7.25	7.26	7.27	7.28	7.29	7.30	7.31	8.1
Expect	0.65	0.63	0.62	0.45	0.35	0.4	0.52	0.56	0.45	0.38
Joy	0.11	0.05	0.03	0.03	0.04	0.05	0.08	0.15	0.17	0.18
Love	0.09	0.06	0.05	0.06	0.06	0.05	0.08	0.09	0.15	0.16
Surprise	0.23	0.52	0.42	0.43	0.35	0.42	0.36	0.35	0.33	0.32
Anxiety	0.63	0.75	0.66	0.65	0.53	0.51	0.48	0.45	0.45	0.44
Sorrow	0.38	0.62	0.65	0.66	0.63	0.68	0.70	0.69	0.65	0.58
Angry	0.21	0.58	0.53	0.75	0.77	0.69	0.68	0.67	0.63	0.56
Hate	0.18	0.52	0.65	0.72	0.73	0.7	0.69	0.68	0.67	0.56

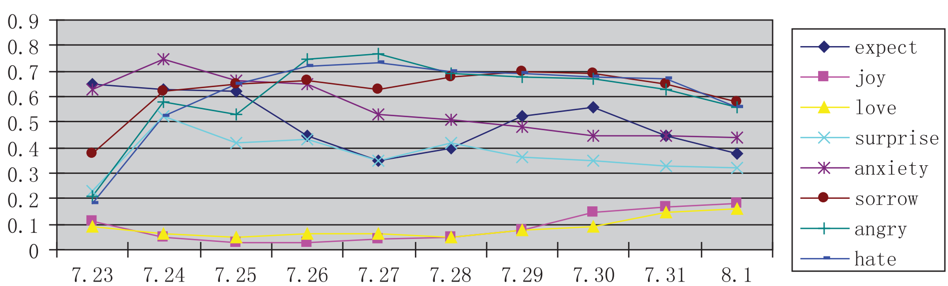

Figure 5. Spark lines for public sentiment after '7.23 Wenzhou Train Collision' on 23 July 2011.

motor car), with the number of deaths increasing, the sadness of the public continue to turn sour, 7.24 night's news conference of the Ministry of Railways should be an important reason for rising resentment and angry of the public.

The third day of '7.23 Wenzhou Train Collision' (7.25): the value of 'angry', 'anxiety', and 'surprise' had declined slightly, mainly due to the central leaders did important instructions that day and the scene of the accident was open again; but the mood of 'hate' was still rising, many Internet users showed strong negative emotion on the manner of disposal of the accidents and China's current social situation.

The fourth day of '7.23 Wenzhou Train Collision' (7.26): the value of 'expect' decreased significantly, because the search and rescue work of electric multiple units (EMU) has been to end, and users hope were also declining for a miracle to happen; the sentiment of 'angry' and 'hate' increased significantly, mainly because publishing of the list of victims and compensation for death irritated nerves of the general public, and tracing the cause of the accident of EMU also made the public angry.

The fifth day of '7.23 Wenzhou Train Collision' (7.27): the value of 'angry' and 'hate' reached the highest value 0.77 and 0.73, respectively; the main reason should be that relevant departments began to cremate the remains of the victims, and the public expressed strong negative emotions.

The sixth day of '7.23 Wenzhou Train Collision' (7.28): the value of 'angry' and 'hate' had declined slightly, the sentiment of 'expect' and 'surprise' rose slightly, refer to the dynamic news that day, Premier Wen Jiabao appeared at the scene of the accident to appease the public.

The seventh day of '7.23 Wenzhou Train Collision' (7.29): the value of 'sorrow' reached the highest value, maybe 'the first seven' caused the public memory of the victims. Several other negative sentiments had decreased, refer to the dynamic news, Premier Wen Jiabao visited to the wounded and increased the death benefits should be important factors.

The eighth day of '7.23 Wenzhou Train Collision' (7.30): the values of positive sentiments such as 'expect', 'joy', and 'love' had risen, several negative sentiments declined, the reason may be that public sentiment began to become dull after a week of hot discussion about emergencies.

The 9th day (7.31) and 10th day (8.1) of '7.23 Wenzhou Train Collision': the public's negative sentiments: 'sorrow', 'angry', and 'hate' were still significantly higher than positive sentiments 'joy' and 'love', but the overall trend moved closer to the middle, the entire incident came into the late soul-searching and processing stage, the public's focus gradually shifted.

5. Conclusion

In this paper, we performed a sentiment analysis of microblogs on '7.23 Wenzhou Train Collision' publicly broadcasted on the Sina Weibo platform between 23 July and 1 August 2011. We regard an individual microblog as a microscopic, temporally authentic instantiation of sentiment, and collected data on both dynamic news events within 10 days after '7.23 Wenzhou Train Collision' and 2,024,792 microblogging posts on the '7.23 Wenzhou Train Collision'; built fuzzy sentiment ontology containing a total of 2090 sentiment entries, combined with the degree words, negative words, and other semantic factors; took into account the influence and timing characteristics of microblogging texts, computed the sentiment of microblogging; and at last we obtained the public's sentiment curve on '7.23 Wenzhou Train Collision' between 23 July and 1 August 2011 and analysed the cause of changes of public sentiment combined with the significant developments in news events. We found that the sentiment expressions of the public for emergencies were closely correlated with the government's handling methods and means for events.

The main contribution of this paper is three-fold: first, we have built fuzzy sentiment ontology based on HowNet and semantic similarity for sentiment analysis in Chinese. Second, we argue that sentiment analysis of minute text corpora (such as microblogging) is efficiently obtained via the approach of ontology and semantic that requires no training or machine learning. The constructed method can also be applied in other fields, such as blog, online reviews in E-commerce, news comments, etc. Finally, we speculate that collective sentiment trends can be modelled and predicted using a large number of online platforms (including microblogging). By analysis of sentiment trends combined with the progress of events, the results provide the necessary reference to control public opinion for the government.

Funding

This work is partially supported by the NSFC [grant number 70971099 and 71371144], the Fundamental Research Funds for the Central Universities [grant number 1200219198], Shanghai philosophy and social science planning projects [grant number 2013BGL004], and the Doctoral Thesis Funding for Soft Science Research from the Shanghai Science and Technology Development Funds [grant number 12692193000].

References

Barbosa, L., & Feng, J. (2012, May). *Robust sentiment detection on Twitter from biased and noisy data*. Proceedings of the 23rd annual conference on computational linguistics (Coling 2010). Retrieved May 3, 2012, from www.aaai.org/ocs/index.php/ICWSM/ICWSM11/paper/view/2857/3251

Berminghan, A., & Smeaton, A. (2012). *Classifying sentiment in microblogs: Is brevity an advantage?* Proceedings of the 19th ACM international conference on information and knowledge management (CIKM'10). Retrieved May 3, 2012, from http://doras.dcu.ie/15663/1/cikm1079-bermingham.pdf

Bollen, J., Gonçalves, B., Ruan, G., & Mao, H. (2011). Happiness is assortative in online social networks. *Artificial Life, 17*(2), 237–251.

Bollen, J., Pepe, A., & Mao, H. (2012, May). Modeling public mood and emotion: Twitter sentiment and socioeconomic phenomena. Retrieved May 3, 2012, from http://arxiv.org/abs/0911.1583

British Develop Public Opinion Analysis Software. (2005, April 11). Global Times (Version 6).

Changqin, Q., & Fuji, R. (2010). A blog emotion corpus for emotional expression analysis in Chinese. *Computer Speech and Language, 24*, 726–749.

Chaovalit, P., & Zhou, L. (2005a). Movie review mining: A comparison between supervised and unsupervised classification approaches. In *Proceedings of the 38th annual Hawaii international conference on system sciences* (pp. 1–9). Hawii.

Chaovalit, P., & Zhou, L. (2005b). Movie review mining: A comparison between supervised and unsupervised classification approaches. In *Proceedings of the 38th annual Hawaii international conference on system sciences* (pp. 112–115). Hawii.

Cheng, Y. (2011, May). *Micro blogs now no 3 online source of info*. China Daily Hong Kong Edition (2011, March 26, p. 03), Hong Kong. Retrieved May 3, 2012, from http://www.chinadaily.com.cn/china/2011-03/26/content_12230674.htm

Cheong, M., & Lee, V. C. S. (2011). A Microblogging-based approach to terrorism informatics: Exploration and chronicling civilian sentiment and response to terrorism events via Twitter. *Information Systems Frontiers*, *13*(1), 45–59.

Collett, M. (2012). China turns to tweeting: Exploring the problematic use of tweeting in China. *International Journal of Business and Social Science*, *3*(2), 91–94.

Dave, K., Lawrence, S., & Pennock, D. M. (2003a). Mining the peanut gallery: Opinion extraction and semantic classification of product reviews. In *Proceedings of the 12th international World Wide Web conference (WWW)* (pp. 519–528). New York, NY: ACM.

Dave, K., Lawrence, S., & Pennock, D. M. (2003b). Mining the peanut gallery: Opinion extraction and semantic classification of product reviews. In *Proceeding of 12th international conference on World Wide Web* (pp. 519–528). Budapest: ACM Press.

Davies, A., & Ghahramani, Z. (2011). *Language-independent Bayesian sentiment mining of Twitter*. Proceedings of the 5th SNA-KDD workshop (SNA-KDD'11). Retrieved May 3, 2012, from http://www.alexdavies.net/wordpress/wp-content/uploads/2011/09/Language-Indepedent-Bayesian-Sentiment-Mining-of-Twitter.pdf

Diakopoulos, N. A., & Shamma, D. A. (2010). Characterizing debate performance via aggregated twitter sentiment. In *Proceedings of the 28th international conference on human factors in computing systems* (pp. 1195–1198). New York: ACM Press.

Ding, L. (2011). Microblogging in network society. *Journal of Digital Research and Publishing*, *18*(2), 1–8. Retrieved May 3, 2012, from http://www.artichokewebdesign.com/ARIN6912/PDFs/Ling_Ding_Microblogging.pdf

Eguchi, K., & Lavrenko, V. (2006). Sentiment retrieval using generative models. In *Proceedings of the 2006 conference on empirical methods in natural language processing* (pp. 345–354). Morristown, NJ: Association for Computational Linguistics.

Ekman, P. (1993). Facial expression and emotion. *American Psychologist, 48*(4), 384–393.
Go, A., Bhayani, R., & Huang, L. (2009). *Twitter sentiment classification using distant supervision* (Technical Report). Stanford: Stanford Digital Library Technologies Project.
Gruzd, A., Doiron, S., & Mai, P. (2011, May). Is happiness contagious online? A case of Twitter and the 2010 Winter Olympics. Retrieved May 3, 2012, from http://ieeexplore.ieee.org/xpl/freeabs_all.jsp?tp=&arnumber=5718715
Hong, X., Hongfei, L., & Yu, P. (2008). Emotional vocabulary ontology construction. *Journal of the China Society for Scientific and Technical Information*, *27*(2), 180–185.
Hou, S., & Zhao, Z. (2011). Research on product market analysis model for microbloggings platform. *Microcomputer Application*, *27*(4), 4–7.
HowNet [R]. (2007). HowNet's home page. Retrieved from http://www.keenage.com
Huang, C. M., Chan, E., & Hyder, A. A. (2010). Web 2.0 and internet social networking: A new tool for disaster management? Lessons from Taiwan. *BMC Medical Informatics & Decision Making*, *10*, 57–61.
Java, A., Song, X., Finin, T., & Tseng, B. (2007). Why we Twitter: Understanding microblogging usage and communities. In *Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 workshop on web mining and social network analysis* (pp. 56–65). New York, NY: University of Maryland, Baltimore County.
Jiang, L., Yu, M., Zhou, M., Liu, X., & Zhao, T. (2011). Target-dependent Twitter sentiment classification. In *Proceedings of ACL* (Vol. 23, pp. 45–56). Stroudsburg: Association for Computational Linguistics.
Kouloumpis, E., Wilson, T., & Moore, J. (2011). *Twitter sentiment analysis: The good the bad and the OMG*. Proceedings of the fifth international AAAI conference on weblogs and social media. Retrieved May 3, 2012, from www.aaai.org/ocs/index.php/ICWSM/ICWSM11/paper/view/2857/3251
Lin, C. d. (2006). Emotional problems of socialism in psychology. *Social Psychology, 21*(83), 37–62.
Liu, B., Hu, M., & Cheng, J. (2005). Opinion observer: Analyzing and comparing opinions on the web. In *Proceedings of the 14th international World Wide Web conference (WWW)* (pp. 342–351). New York, NY: ACM.
Liu, Q., & Li, S. (2002). *Semantic similarity computing based on HowNet* (The Third Chinese Lexical Semantics Workshop). Taipei.
Liu, S., Björkstén, J., & Lagerdahl, C. (2012). *Best practices in Chinese microblog communications* (Eastwei MSL Executive Whitepaper). Retrieved May 3, 2012, from http://www.eastweimsl.com/Eng_Best_Practice_Microblogging_Eastwei_MSL_Executive_Whitepaper.pdf
Mishne, G., & Glance, N. (2006). *Predicting movie sales from blogger sentiment*. AAAI 2006 Spring symposium on computational approaches to analysing weblogs, New York.
Na, Ri-sa, Liu, Y., & Li, Y. (2010). Semantic fuzzy calculation and product recommendation based on online reviews. *Journal of Guangxi Normal University: Natural Science Edition*, *28*(1), 143–146.
Naaman, M., Boase, J., & Lai, C. H. (2010). *Is it all about me? User content in social awareness streams*. Proceedings of the 2010 ACM conference on computer supported cooperative work, Savannah.
Osgood, C. E., & Tannenbaum, P. H. (1955). The principle of congruity in the prediction of attitude change. *Psychological Review, 62*(1), 42–55.
Oxford Dictionaries Online. (2012). Microblogging. Retrieved May 3, 2012, from http://oxforddictionaries.com/definition/microblogging
Pan, Y., & Lin, H.-f. (2008). Restaurant reviews mining based on semantic polarity analysis. *Computer Engineering, 34*(17), 208–210.
Pang, B., Lee, L., & Vaithyanathan, S. (2002). Thumbs up? Sentiment classification using machine learning techniques. In *Proceedings of the 2002 conference on empirical methods in natural language processing (EMNLP)* (pp. 79–86). Morristown, NJ: Association for Computational Linguistics.
Pfeiffer, H. D., & Tourte, G. (2012). Twitter, information sharing and the London Riots? *Bulletin of the American Society for Information Science and Technology*, *38*(2), 49–57.
Qu, Y., Huang, C., Zhang, P., & Zhang, J. (2011). Microblogging after a major disaster in China: A case study of the 2010 Yushu Earthquake. In *Proceedings of the 2011 ACM conference on computer supported cooperative work CSCW 2011* (pp. 25–34). New York: ACM Press.
Sina News Center. (2012). Sina microblogging registered users exceeded 300 million. Retrieved May 3, 2012, from http://news.sina.com.cn/m/news/roll/2012-03-31/003724202903.shtml
Sparck Jones, K. (1965, June and October). Experiments in semantic classification. *Mechanical Translation and Computational Linguistics*, *8*(3 and 4), 25–43.
Stone, P. J., Dunphy, D. C., & Smith, M. S. (1966). The general inquirer: A computer approach to content analysis. PsycINFO, Oxford: M.I.T. Press, p. 651.
Subasic, P., & Huettner, A. (2001, August). Affect analysis of text using fuzzy semantic typing. *IEEE Transactions on Fuzzy Systems, 9*(4), 32–36.
Thelwall, M., Buckley, K., & Paltoglou, G. (2011). Sentiment in Twitter events. *Journal of the American Society for Information Science and Technology*, *62*(2), 406–418.
Turney, P. (2002, July 8–10). Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews. In *Proceedings of the 40th annual meeting of the Association for Computational Linguistics (ACL)* (NRC 44946, pp. 417–424). Philadelphia, PA: Computing Research Repository.
Wang, H., Yin, P., Yao, J., & Liu, J. N. K. (2012). Text feature selection for sentiment classification of Chinese online reviews. *Journal of Experimental & Theoretical Artificial Intelligence*. doi:10.1080/0952813X.2012.721139, 2012: 1–15, iFirst.
Xu, L., & Ling, H. (2007). *The text affective computing based on semantic resource*. Dalian: Dalian University of Technology Computer Science and Engineering.

Xu, X., & Tao, J. (2003). Study on affective division in Chinese emotion system. In *The 1st Chinese conference on affective computing and intelligent interaction* (pp. 199–205). Beijing.

Yan, J., Bracewell, D. B., Ren, F., & Kuroiwa, S. (2008). The creation of a Chinese emotion ontology based on HowNet. *Engineering Letters, 16*(1), EL_16_1_24.

Yin, P., Wang, H., & Guo, K. (2013). Feature-opinion pair identification of product reviews in Chinese: A domain ontology modelling method. *New Review of Hypermedia and Multimedia*, *19*(1), 3–24. doi:10.1080/13614568.2013.766266.

Yu, L., Asur, S., & Huberman, B. A. (2011a). *What trends in Chinese social media?* Proceedings of the 5th SNA-KDD Workshop (SNA-KDD'11). Retrieved May 3, 2012, from http://www.hpl.hp.com/research/scl/papers/chinatrends/china_trends.pdf

Yu, L., Asur, S., & Huberman, B. A. (2011b). Artificial inflation: The true story of trends in Sina Weibo. Retrieved May 3, 2012, from http://www.hpl.hp.com/research/scl/papers/chinatrends/weibospam.pdf

Zhang, L., Ghosh, R., Dekhil, M., Hsu, M., & Liu, B. (2011). *Combining lexicon-based and learning-based methods for Twitter sentiment analysis* (Hewlett-Packard Labs Technical Report, HPL-2011-89).

Zhang, M., Sun, C., & Liu, W. (2011). *Identifying influential users of microbloggingging services: A dynamic action-based network approach*. PACIS 2011 Proceedings, Brisbane, Australia. Retrieved May 3, 2012, from http://aisel.aisnet.org/pacis2011/223/

Zhang, Y., Li, Z. M., Ren, F. J., Li, Z., & Kuroiwa, S. (2005). Semi-automatic emotion recognition from textual input based on the constructed emotion thesaurus. In *International conference on natural language processing and knowledge engineering* (pp. 571–576). Wuhan: IEEE Xplore.

Zheng, L., Ren, Y., Liu, X., & Xu, H. (2011). S*tudying the management mechanism of government microblogs in China: A research plan*. Proceedings of the 5th international conference on theory and practice of electronic governance (ICEGOV '11). Retrieved May 3, 2012, from http://www.sirpa.fudan.edu.cn/picture/article/56/41/24/d3693a5d4527987ecaa870bed428/46a51192-63c2-47a2-82c3-60629d903f4d.pdf