*International Journal of Knowledge and Language Processing* *www.ijklp.org*
KLP International ⓒ2015 ISSN 2191-2734
Volume 6, Number 3, **2015** pp.1—10

# A Method of Constructing on Micro-Blog Content Credibility Model

Chai Jing[1], Zhang Yangsen[2], Liu Qingsong[3]

[1]Institute of intelligent information processing.
Beijing Information Science and Technology University
Beijing, 100192
handsomecj@126.com

[2]Institute of intelligent information processing.
Beijing Information Science and Technology University
Beijing, 100192
zhangyangsen@163.com

[3]Institute of intelligent information processing.
Beijing Information Science and Technology University
Beijing, 100192
liuqingsong@163.com

Selected Paper from Chinese Lexical Semantic Workshop 2015

ABSTRACT. *This study researched from the direction of the index system in statistics and the idea of machine learning classification algorithm respectively, and the goal was to research methods and technology of accurate judgment of micro-blog information credibility. Taking aim at the uniqueness of the structure and content of micro-blog short text, content credibility model of micro-blog was constructed. In this paper, micro-blog content credibility model was constructed by combining multi-classifier, firstly, identifying the polarity tendency of word; Secondly, extracting the single words and the dual words of micro-blog text as the feature set of Bayesian Emotion Classification; Finally, using the results of the Bayesian Emotion Classification and other seven rumors micro-blog classification features as the feature set of Support Vector Machine(SVM) classifier, and the classification results could determine whether each micro-blog was rumors or not, and the content credibility model of micro-blog was constructed completely. Experimental results showed the effectiveness of this method.*

**Key words:** Micro-blog; Content Credibility; Multi classifier

1. **Introduction.** The rapid development of Information Technology (IT) promoted the popularity and usage of the Internet. The virtual community [1-2] is the phenomenon of social agglomeration derived from the Internet, which evolve from the early Blog and BBS to the now popular Micro-blog. According to China Internet Network Information Center (CNNIC) [3], as of June 2014, the number of China's Internet users reached 632 million, and the total number of mobile phone users increased to 527 million. Among huge amount of China Internet users, Micro-blog, a popular social network, provided a good communication platform for Chinese users.

Micro-blog is a kind of integrated and open social service platform. With the characteristics of real-time, brief, agility, shortcut, micro-blog has quickly become an important medium for millions of Internet users to publish and obtain social information. Micro-blog [4-5] not only bring great convenience for the information communication and transmission, but also provide an opportunity for the spread of false information (Internet rumors [6]), and most micro-blog users can't tell the authenticity of micro-blog. Internet rumors play an important role in the public emergencies. Micro-blog, as a grassroots medium, give the spread of Internet rumors a 24-hour "power engine", boost the formation and spreading of rumors, and cause the spread of false information. So a large number of micro-blog rumors appear on the micro-blog platform. The so-called micro-blog rumors refer to the micro-blog which is difficultly distinguished with the public normal level of thinking and analysis ability and its content is unsubstantiated. Rumors [7] not only misdirect netizens, seriously disturb the normal order of network transmission, directly threat to the healthy development of the Internet of the present and the future, but also reduce the credibility of network transmission, pollute media environment, become the "Trojan horse" of network transmission. Therefore, the research on the Chinese micro-blog information credibility is an urgent issue to be solved at the moment, and it is also the hot spot of researchers. Especially in the critical period of information society transition of our country, how to judge and identify the false information, select and use the micro-blog information of high credibility and high value, guarantee the healthy and orderly development of the Internet, has become a problem that drew a lot attention from enterprises and government.

2. **Preprocessing of micro-blog content.**

2.1. **Discrimination of polarity of emotional words.** This study identified the polarity of emotion words by using HowNet and the emotion words ontology together. This project consisted of three parts:

(1) Selecting benchmark word pair. The benchmark word pair [8-9] was emotion words set having obvious emotional tendency, strong emotion and well representativeness. The selection of benchmark word set must meet the following conditions: high emotion strength, well representativeness, wide coverage, not redundant in terms of semantics. The emotion words ontology was divided into 21 categories, about four or five emotion words were

picked from each category, and these words had great emotion strength and no auxiliary emotion. The selected emotion words were searched in Google, and the words having big return results were selected. And then the HowNet knowledge base was used to replace and fuse the words having same or similar semantic meaning. Finally, Su-Ge Wang [10]'s benchmark word set was used.

(2) Obtaining the words set of equivalent emotional tendency based on HowNet sememe limiting condition. The stronger the semantic similarity between words was, the closer the emotional tendency was. While the calculation of words similarity returned to the calculation of sememe similarity, and the structure description of sememe by using Knowledge Database Mark-up Language (KDML) would be involved

ulteriorly. The degree of semantic similarity between different sememes depended on the number of equal sememe and sememe categories, so the equivalent emotional tendency words set SWL, could be got by limiting conditions. The basic sememe must be equal in the set, that was the sememes, at the same level in the tree hierarchy, at least half of the subprime sememes and relation sememes were equal, and if sememes contained symbol sememes, at least one symbol sememe was equal, otherwise this condition didn't be considered.

(3) Calculating the lexical semantic tendency. The numerical range of computing results of lexical semantic emotional tendency was [-1, 1]. The similarity calculation of SWL and benchmark words would be converted to the similarity calculation of words and benchmark words. The semantic similarity maximum was introduced in Yang Yubing [11], which significantly improved the accuracy. The formula was as follows:

$$Orientation(W) = \left(\frac{1}{\alpha}\sum_{i=1}^{k} Simlarity(key-p_i, W) + \frac{1}{\beta}\max_{i=1\ldots k} Simlarity(key-p_i, W)\right) - \left(\frac{1}{\alpha}\sum_{j=1}^{k} Simlarity(key-n_j, W) + \frac{1}{\beta}\max_{j=1\ldots k} Simlarity(key-n_j, W)\right)$$ Formula (2-1)

In above formula, k represented k pairs benchmark words. The benchmark word could be divided into commendatory benchmark words and derogatory benchmark words, and $\mathrm{key}-\mathrm{p_i}$ represented the commendatory benchmark word, $\mathrm{key}-\mathrm{n_j}$ represented the derogatory benchmark words, $\mathrm{Simlarity(w1, w2)}$ represented the semantic similarity of words, $\alpha$、$\beta$ was the adjustable parameters. Then the similarity of all elements in the set with benchmark words was calculated:

$$OrientationSet(SWL) = \frac{\sum_{i=1}^{len1} Orientation(W_i)}{len1}$$ Formula (2-2)

Where len1 was the total number of words in the WordList1, and the result was the semantic emotion tendency value of undetermined emotion words. The emotion polarity of those words was determined by the formula (2-4):

$$Orientation(word) = OrientationSet(SWL)$$ Formula (2-3)

$$Judge\ (word) = \begin{cases} 1 & 0.1 < Orientation(word) \leq 1 \\ 0 & -0.1 \leq Orientation(word) \leq 0.1 \\ -1 & -1 \leq Orientation(word) < -0.1 \end{cases} \quad \text{Formula (2-4)}$$

If the value of Judge (word) was 1, we could think the word was positive. If the value of Judge (word) was 0, that the word had no emotion or its emotional tendency was so weak that it could be ignored. If the value of Judge (word) was -1, we could think the word was negative in terms of emotion.

2.2. **Extraction and calculation of micro-blog content features.** The preprocessing of micro-blog content was actually the process of extracting emotion classification features and rumors classification features. Due to the particularity of micro-blog—short text structure and the meaning expressed by content is fuzzy, general features extraction methods are not suitable for the features extraction of micro-blog in the direction of credibility analysis. For example, in this paper, the false information was classified by combining chi-square test with the mutual information as features. In the results, the total accuracy was only 35.33%, and the recognition rate of rumors was 37.45%, and the correct recognition rate of real micro-blog was 34.57%, because the features selection was not in conformity with the actual situation, most of the rumors and real micro-blog could not be classified successfully. Considering the particularity of micro-blog and purpose of identifying the real micro-blog, the micro-blog content credibility model was built by selecting eight features. These features and their calculation methods were:

(1) Emotional tendency. Each micro-blog attach emotional tendency of users more or less, and the rumor micro-blog attach potential information having strong emotional tendencies. Micro-blog was classified by using the Naïve Bayesian classifier, and 0 or 1 represented different categories, 0 represented the negative emotional tendency, 1 represented the positive emotional tendency.

(2) The number of URL. Micro-blog users want to prove the authenticity of their points and hope to get the support of others by referring URLs. In the rumors micro-blog, URL appears frequently, rumormongers hope that the rumor has "real credibility" for other users. Due to the one-to-one corresponding relations of "http://" and URL, the number of URL could be obtained by calculating "http://".

(3) The number of "@". As a unique label element, "@" has been widely applied to the rumor micro-blog, rumor mongers rely on others to prove the authenticity of their rumors. We could directly calculate the number of "@" occurrences.

(4) The number of topics. As a unique label element, topics appear with symbol forms of "#" or "【】". In some common rumor micro-blog, the rumor mongers use the hot topics to deceive people. If the number of "#" was M and the number of "【】" was N, the sum of M and N was the number of topics.

(5) The number of emotion words. First of all, the emotion vocabulary ontology knowledge base compiled by Dalian University of Technology was used to build the emotion dictionary. Then, Zhang Huaping word segmentation system was used to split

micro-blog content. Finally, the text which had been processed was traversed and it was matched against the emotion words, we could get the number of emotion words.

(6) Emotion icon. They are used to express users' excitement, scorn and dismay. For example,  shows that user is very excited,  shows that user is sad. By tailing after rumors, we could find that the number feature of entirety emotion icon was special when a lot of rumors gathered together. The most of emotion icon in rumors were some specific emotion icon in the emotion icon dictionary, which had a small number or gather together. When their features were calculated, first of all, the sina micro-blog emotion icon dictionary was constructed. Then, "[]" was matched against regular expression, and the words in "[]" were extracted, and the number of corresponding emotion icon plus 1 if the word was in the emotion icon dictionary. Finally, the number of emotion icon was returned.

(7) The number of book number. In a large number of rumors, a large portion of rumor mongers will use specific reports, magazines, masterwork and film to spread rumors. Rumor mongers use the authority of books to let people believe the authenticity of rumors. Each micro-blog was pattern matched against regular expression, and the number of "《》" was selected and calculated, and the result was returned at last.

(8) Micro-blog is original or not. We found out that the possibility of the original micro-blog was rumor micro-blog was higher than user forwarded micro-blog by analyzing micro-blog data.

3. **Micro-blog content credibility model.** Constructing the micro-blog content credibility model was actually evaluating the credibility of every micro-blog posted by a user; this was a dichotomy result—either a rumor or an authentic micro-blog. The construction idea was as follow:

Constructing the rumor classifier based on SVM by using the micro-blog content features described in section 2.2 to classify all micro-blog and to have one to one correspondence from user to user's micro-blog, to construct the micro-blog content credibility model. This model could not only identify and distinguish rumors of the existing user, but also could deal with rumors of users unfamiliar and from other social network, among them, there were some ads. The content of advertising micro-blog may be true or false, so advertising micro-blog may be authentic micro-blog or rumor micro-blog. For all kinds of ads in micro-blog, the micro-blog content credibility model constructed in this paper could distinguish the true and false of ads.

3.1. **Micro-blog text emotion classification based on Bayesian classifier.** Micro-blog emotion classification was the basis of construction of the micro-blog content credibility model. As a feature of rumor classifier based on SVM, the accuracy of classification result directly affected the results of SVM classification. Before using Bayesian classifier for emotion classification, we needed to preprocess micro-blog content, then to classify them one-time and arrange the results. The steps of Bayesian classification were as follows.

Step1: Preprocessed samples of training sets and test sets, which was removing the stop words in the text.

Step2: Separated the text after preprocessing, and calculated features of all texts.

Step3: Used Bayesian classifier to train the training sets, and constructed the emotion classifier model, and tested the test sets, and predicted the emotion categories of texts in test set.

Step4: The corpus with the best classification results was used as the training set of Bayesian classification. Micro-blog collected was use as the test set to classify emotion. Finally, the results were used as a feature of SVM rumor classification.

The results of emotion classification by using Bayesian classifier were used as one of features of SVM. Semantic polarity tendency provided a unified interface to determine whether the word had emotion or not.

3.2. **Evaluation of micro-blog content credibility based on SVM.** SVM classifier was a kind of learning method based on structural risk minimization norm. This method selected a group support vector (SV) from training sets, which made the partition be equal and classified the whole data set. The process of selection SV was actually the process of construction model. Then the model was used to classify test set. SVM classification was performed mainly from four aspects respectively:

(1) Prepared data set, and converted the data to formats supported by a library for SVM (LIBSVM). The data were rumor micro-blog and real micro-blog collected, the rumor micro-blog not determined would be removed. In this paper, eight features were extracted for SVM classification, the feature value of each feature was calculated by using method of each feature, and subsequently feature vector was generated. The vector dimension was eight, and the value of vector was the value of each feature, as shown in FIGURE 3.1.

| nikename | emotion classification | number of URL | number of "@" | number of topic | emotion icon | book number | original? |
|---|---|---|---|---|---|---|---|
| xiaotudou | 0 | 0 | 0 | 0 | 1 | 0 | 1 |
| muyu | 0 | 1 | 2 | 3 | 1 | 1 | 0 |
| sanyuedemaomi | 1 | 3 | 1 | 0 | 1 | 0 | 0 |
| xiena | 1 | 1 | 0 | 0 | 1 | 1 | 1 |
| xiena | 0 | 0 | 2 | 1 | 3 | 2 | 1 |

FIGURE 3.1. SKETCH MAP OF FEATURE

(2) Scaled all the data. The "svmscale" was used to scale the original sample data, the scaling ranges were in [-1, 1]. The purpose of scaling was to speed up the calculation.

(3) Tuned parameters. The goal of tuning parameters was to construct the classification model, which could obtain the best results. The obtained model was used by SVM system.

(4) The test set was classified by using the SVM training model. The classification results were 0 or 1. 1 represented that it was not rumor micro-blog, and 0 represented rumor micro-blog.

After rumor micro-blog was classified by using SVM classifier, each user's micro-blog could be determined whether it was a rumor. Micro-blog user _Mr_Namo published 13 micro-blogs, the credibility of those micro-blogs were shown in FIGURE 3.2.

| Sequence number | nickname | SVM rumor identification |
|---|---|---|
| 1 | _Mr_Namo | 0 |
| 2 | _Mr_Namo | 0 |
| 3 | _Mr_Namo | 0 |
| 4 | _Mr_Namo | 0 |
| 5 | _Mr_Namo | 0 |
| 6 | _Mr_Namo | 0 |
| 7 | _Mr_Namo | 0 |
| 8 | _Mr_Namo | 0 |
| 9 | _Mr_Namo | 0 |
| 10 | _Mr_Namo | 0 |
| 11 | _Mr_Namo | 0 |
| 12 | _Mr_Namo | 1 |
| 13 | _Mr_Namo | 0 |

FIGURE 3.2. USER'S PARTIAL MICRO-BLOG CONTENT CREDIBILITY

4. **Experiment results and analysis.**

4.1. **Classification evaluation method.** Generally, three performance measures were used to evaluate the results of classification. They were:

(1) Precision

$$P = \frac{\text{sample predicted with the model accurately}}{\text{sample in all test sets}} \times 100$$

*Formula (4-1)*

(2) Recall

$$R = \frac{\frac{\text{positive sample predicted with the model accurately}}{\text{positive sample in test set}} + \frac{\text{negative sample predicted with the model accurately}}{\text{negative sample in test set}}}{2}$$

Formula (4-2)

(3) F value

$$F = 2 * P * \frac{R}{P+R} \times 100\%$$

Formula (4-3)

4.2. **Bayesian emotion classification results assessment.** Because the training set and test set used by Bayesian classifier were different, the results of classification were also different. This paper selected 21257 micro-blogs of the 2013 NLP&CC micro-blog mood sample corpus, and these micro-blogs were marked with Data1. We selected 2215 micro-blogs collected, and marked them with Data2. We selected 32186 micro-blogs which were lab resources corpus, and marked them with Data3. Data1 and Data3 were training set, and Data2 was test set. The combination and distribution of all data as shown in TABLE 4.1 and TABLE 4.2.

TABLE 4.1. BAYESIAN DATA DISPLAY LIST

| *Classifier symbol* | *Training set* | *Number of micro-blog* | *Test set* | *Number of micro-blog* |
|---|---|---|---|---|
| *Bayes1* | *NLP&CC micro-blog mood sample* | *21257* | *Manually annotated corpus* | *2215* |
| *Bayes2* | *Lab text emotion sample* | *32186* | *Manually annotated corpus* | *2215* |
| *Bayes3* | *NLP&CC micro-blog and lab data* | *53443* | *Manually annotated corpus* | *2215* |

TABLE 4.2. BAYESIAN CLASSIFICATION RESULT LIST

| *Classifier symbol* | *Precision* | *Recall* | *F value* |
|---|---|---|---|
| *Bayes1* | *0.713* | *0.735* | *0.724* |
| *Bayes2* | *0.524* | *0.701* | *0.600* |
| *Bayes3* | *0.605* | *0.710* | *0.653* |

Under the circumstance of same feature extraction methods, we could see that the result of Bayes1 was the best, so the NLP&CC micro-blog mood sample was selected as the final training set. All micro-blog texts crawled were as test set. The combined corpus was classified with Bayesian classifier; the results was as a feature of SVM classifier.

4.3. **SVM rumor classification result assessment.** TABLE 4.3 was the predicted result of all test set through adjusting the SVM parameters and using the best training model. TABLE 4.4 was the classification result of all micro-blogs as test set.

TABLE 4.3. SVM CLASSIFICATION RESULT

| *Classifier symbol* | *Precision* | *Recall* | *F value* |
|---|---|---|---|
| *SVM1* | *0.785* | *0.772* | *0.778* |

TABLE 4.4. SVM CLASSIFICATION FINAL RESULT

| *Test corpus identification* | *Precision* | *Recall* | *F value* |
|---|---|---|---|
| *Rumor identification* | *0.949* | *0.975* | *0.962* |
| *Normal micro-blog identification* | *0.725* | *0.763* | *0.744* |

The statistical results of table 4.4 showed that the evaluation index of rumor was higher than the normal micro-blog. Analysis of reasons were as follows:

1) The number of normal micro-blog was larger than the rumor, and the collection efficiency of rumor micro-blog was low.

2) Due to the SVM rumor classification model extracted rumor features, the amount of classification information obtained by SVM identifying rumor was larger than the normal micro-blog.

5. **Conclusions.** Entering the Web 2.0 era, user published micro-blog not only costs low price, but also needs not bear any legal responsibility, which causes all sorts of false information has overflowed, and the information credibility problem is also on the agenda, and the evaluation of micro-blog credibility has become an urgent need, and the calculation of information credibility has drawn more and more attention. In this paper, the double classifiers nested scheme was put forward to complete the construction of micro-blog content credibility model by analyzing the advantages and disadvantages of analysis and mining methods about the credibility evaluation of micro-blog short text. The model was based on statistics and machine learning algorithms. The experiment result illustrated that this method achieved accurate evaluation of micro-blog information credibility. The reasonable selection of evaluation measure and the use of double classifiers ensured the credibility and theoretical rationality of this model results.

At present, micro-blog has been a gateway of gaining mass data and mining potential information value, and an increasing number of scholars invest a lot of effort in it. The research of this study was very valuable, but this system limited to time and environment factors had some shortcomings. The next step work is mainly from the following aspects: (1) Function expansion of micro-blog data acquisition system. We can collect data by using the multi-user multithreading. (2) Construction of micro-blog user credibility model. If the user has the high authority, other users will trust him. The measurement features of user authority include fan level, focus level, VIP and so on. Combining micro-blog user credibility model with micro-blog content credibility model can get a better effect. (3) Combined with the theory of communication. It's worth noting that is not more technology is used and more ambits are involved the effect of the Chinese micro-blog credibility evaluation system is better. The key is that selecting the right object can calculate the satisfactory and convincing results. In recent years, researchers found that the rumors and normal micro-blog have lots differences in the communication process. The topological structure of micro-blog communication was constructed by micro-blog communication depth and comments and forwarding reasons to extract features. Then micro-blog information credibility model based on the communication theory was constructed by using the machine learning approach.

## REFERENCE


[1] Zhao Pengwei, Ma Lin. Research Overview of Virtual Community. Theory and Practice of Intelligence. 2013.36(7): 119-123

[2] Saaed Valadbaygi. Virtual Community. Index on Censorship. 2011.40(1):133-137

[3] CNNIC. The 34th China Internet Network Development State Statistic Report.2014. http://www.cnnic.net.cn/

[4] Zhang Siqi. The Micro-blog Culture Research. Changchun: Jilin University. 2012

[5] Zhang Lin. Research of Microblog's Development in China. Nanchang: Jiangxi University of Finance & Economics. 2012

[6] Ren Yiqi, Wang Yalei. Research on the Evolution Mechanism of Micro-blog Rumors.Journal of Intelligence. 2012.31(5): 50-54

[7] Chen Qiang, Zeng Runxi. Micro-blog Rumors: Concept, Function and Govern. 2013.1: 2-7

[8] Chen Jianmei, Lin Hongfei, Yang Zhihao. Automatic acquisition of emotional vocabulary based on syntax. CAAI Transactions on Intelligent Systems. 2009. 4(2): 100-106

[9] Peter.D. Turney, Michael. L.Littman. Measuring praise and criticism: Inference of semantic orientation from association .ACM Transactions on Information Systems 2003.21(4):315-346

[10] Wang Suge, Li Deyu, Wei Yingjie et al. A Synonym Based Word Emotion Orientation Discriminating. Journal of Chinese Information Processing. 2009.23(5): 68-74

[11] Yang Yubing, Wu Xianwei. Improved Lexical semantic tendentiousness recognition computing. Computer Engineering and Applications. 2009.45(21): 91-931

[12] Qazvinian V,Rosengren E,Radev D R,et al.Rumor has it:Identifying misinformation in microblogs[C]//Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing.Edinburgh:ACL,2011:1589－1599.