

RESEARCH

Open Access



A web sentiment analysis method on fuzzy clustering for mobile social media users

Li Yang*, Xinyu Geng and Haode Liao

Abstract

Micro-blog has become an emerging application in the Internet in recent years, and affective computing and sentiment analysis for micro-blog have been a vital research project in computer science, natural linguistics, psychology of human, and other social computing fields. In this paper, firstly, fuzzy clustering theory was introduced and source database for micro-blog was constructed. Then, word similarity computation method based on basic emotion word set of HowNet was used to calculate weights of micro-blog emotion words, and micro-blog emotional lexicon was built. Next, using calculation methods for appropriate sentiment value, the whole micro-blog message's emotional values were obtained. Finally, sentiment values from users in different time periods were selected as original data matrix, using the fuzzy clustering algorithm. The users were classified dynamically; meanwhile, dynamic clustering figure was generated. The best classification was obtained by using F statistics test method, and emotion trend graph was predicted from classification results, to more intuitively analyze emotion changes of user. In this paper, using micro-blog information with affective computing, governments, businesses, or enterprises can get different classification results according to the different needs and take the appropriate measures.

Keywords: Fuzzy clustering, Sentiment analysis, Micro-blog, Sentiment intensity

1 Introduction

Since human entered the twenty-first century, people continue to create high-tech product and enjoy various emotions they bring. Internet products all the time burst with its unique personality and creativity; a few years ago, people always soaked in social networking sites such as “kaixin” and “renren”, to share with the working and living of classmates, colleagues, or friends. Nowadays, “micro-blog” is widely used in the daily life of Internet users and plays an important role. In recent years, with the rapid development of Web technologies and the growing and expanding of website, users have become the browser of website contents and the creator of website contents. All kinds of social network service (SNS), such as Facebook, forum, friends-net, renren, Sina micro-blog, and Tencent zone, have become the indispensable stage for globe internet users to show their emotions, communicate with each other, and share information, as well as become windows to get information, show themselves,

and promote marketing. With the development of mobile networks, users can express their personal mood at any time, describe their experience, and elaborate the subjective and objective view of things on the Internet [1].

Micro-blog, as a new type of media, is developing rapidly. It has the characteristics of fast propagation speed, real-time information updating, and strong interaction and so on and has a tremendous impact on social life of Internet users and even the non-users. Micro-blog users can use the computer, mobile devices (such as mobile phones and iPad) to release and share information to the community in a variety of forms through the Internet. On the one hand, users can post messages on micro-blog. On the other hand, users can also find out news of current events, hot topics, reviews of others on the micro-blog, expand their social circle, and make like-minded friends. The man who designed micro-blog platform at first is, Evan Williams, the founder of blog technology, pioneer blogger. In March 2006, he launched a social network platform and micro-blog service that is now known abroad, Twitter. In China, micro-blog, the new term, has become the world's most popular word and two most

* Correspondence: scxdyl@126.com
School of Computer Science, Southwest Petroleum University, Chengdu, China

popular micro-blog platforms are Sina micro-blog and Tencent micro-blog [2].

Nowadays, the social network has become an important part of people's lives; you can talk about something in the QQ space or micro-blog, share the mood like a diary, and record all kinds of emotional color in the daily life. Therefore, Web text data which contains the rich emotion of Internet users expand at an amazing speed, so the analysis of an effective emotional analysis model has become one of the important topics of network social computing.

Micro-blog sentiment computing and analysis includes two aspects: one is to analyze the personal emotion fluctuation in different period of time; the other is to analyze emotions of a group people in different time. The meanings are as following: first, through the establishment of personal emotion charts, emotional states of Internet users in a certain period of time can be analyzed so that decision makers can understand the emotion changes related to Internet users recently and timely can detect abnormal changes of mood, in order to make the corresponding decision to create a healthy and harmonious network society; secondly, through the research on group emotion analysis, emotion dynamic fuzzy clustering graph can be formed by fuzzy clustering analysis on members in group. And we can set different conditions to get different classification results, in order to find your friends. So, the establishment of various kinds of social relations can provide great help to users in life or in business decisions.

In the paper, fuzzy clustering algorithm is applied to micro-blog sentiment analysis and dynamic clustering graph is generated; the best classification is obtained by using F statistics test method, and emotion trend graph is predicted from classification results by SPSS, to achieve visualization on people's emotions.

The remainder of this paper is organized as follows. Section 2 describes the related work on web sentiment analysis methods. Section 3 gives computation method for the sentiment value of sentiment phrase with modifier, as well as sentiment intensity analysis of micro-blog content. Section 4 presents sentiment analysis for QQ zone users. In Section 5, sentiment clustering analysis of micro-blog users on fuzzy clustering algorithm is put forward. Conclusions are summarized in Section 6.

2 Related works

Wang et al. [3] thought that sentiment analysis is to analyze users' comments on the Web, so as to identify the implicit emotional information, and found out the rules of user emotion evolution; Zhao et al. [4] believed that sentiment analysis was called opinion mining, an analysis, processing, induction, and deduction process of subjectivity text with emotional color. Sentiment analysis

involved various difficult and huge challenging researches. Existing research fruits on micro-blog sentiment analysis are listed as follows:

In 2009, Alec Go, Richa Bhayani, and Lei Huang used machine learning method for the first time to analyze the sentiment on the micro-blog text and joined the emoticon sign in the system so as to greatly improve the accuracy of the system [5]. In 2010, according to statistics on written words of product view and opinion polls of customer between 2008 and 2009, Brendan, Ramnath Balasubramanyan, and Bryan R. Routledge concluded that sentiment words used in product view and opinion polls had a high correlation [6]. Alexander Pak and Patrick Paroubek collected data and trained classifier to make research on sentiment analysis and opinion mining on micro-blog [7]; Albert Bifet and Eibe Frank analyzed the difficulties in micro-blog data mining, solved the problem of unbalanced data classes by setting a kappa test sliding window, and achieved good results [8]; Dmitry Davidov, Oren Tsur, and Ari Rappoport built an emotional classification system by supervising learning based on Twitter with common labels, emotion labels. The method did not need too much manual annotation [9]; Cindy Xide Lin, Bo Zhao, and Qiaozhu Mei discussed social network event topic by using the statistical method [10]; Asli Celikyilmaz, Dilek Hakkani, and Junlan Feng classified Twitter text into polar and nonpolar text by probability and statistics mode and then used the sentiment vocabulary to categorize sentiment polarity of Twitter text [11]; Luciano Barbosa and Junlan Feng detected sentiment information of Twitter text by analyzing text structure and text keywords information of Twitter [12-13]. Long Jiang, Mo Yu, and Ming Zhou improved the accuracy of sentiment classification by merging characteristics and adding relevant text [14]. Lei Zhang, Riddhiman Ghosh, and Mohamed Dekhil adopted sentiment dictionary and machine learning methods to classify emotional polarity text in Twitter based on entity in the text [15].

In online videos, Hoque et al. developed an automated system to distinguish between naturally occurring spontaneous smiles under frustrating and delightful stimuli by exploring their temporal patterns given video of both, evaluated and compared two variants of Support Vector Machine (SVM), Hidden Markov Models (HMM), and Hidden-State Conditional Random Fields (HCRF) for binary classification. McDuff et al. presented results validating a novel framework for collecting and analyzing facial responses to media content over the Internet. The framework, data collected, and analysis demonstrated an ecologically valid method for unobtrusive evaluation of facial responses to media content [16, 17].

In terms of opinion mining, a fuzzy propagation modeling for opinion mining by sentiment analysis of online social networks was proposed by Trung et al. A practical

system, called TweetScope, has been implemented to efficiently collect and analyze all possible tweets from customers [18]. S Venkatesh, D Phung, D Bo, and M Berk used machine learning and statistical methods to discriminate online messages between depression and control communities using mood, psycholinguistic processes, and content topics extracted from the posts generated by members of these communities [19]. To understand customers' opinion and subjectivity, a fuzzy propagation modeling for opinion mining by sentiment analysis of on-line social networks was presented and a practical system, called TweetScope, was implemented to efficiently collect and analyze all possible tweets from customers [20, 21]. A crowdsourcing-based method is proposed by Xu et al. [31–35] to process mobile social media data. To improve sentiment analysis consequences, Xing Wu and Shaojian Zhu investigated the limitations of traditional sentiment analysis approaches and proposed an effective Chinese sentiment analysis approach based on emotion degree lexicon to obtain significant improvement in Chinese text [22]. To address these various limitations and issues in non-English languages, in 2014, Wang et al proposed a new scheme to derive dominant valence as well as prominent positive and negative emotions by using an adaptive fuzzy inference method (FIM) with linguistics processors to minimize semantic ambiguity as well as multi-source lexicon integration and development [23].

To solve the difference among affect, feelings, emotions, sentiments, and opinions in text, M. D. Munezero et al. clarified the differences between these five subjective terms and reveals significant concepts to the computational linguistics community for their effective detection and processing in text [24]. A new affective-behavioral-cognitive (ABC) framework was presented to measure the usual cognitive self-report information and behavioral information, together with affective information while a customer makes repeated selections in a random-outcome two-option decision task to obtain their preferred product [25]. Santos-Sanchez et al. proposed a novel system to obtain data of interest from a Web search engine by analyzing the emotional and sentimental content of each page, which made possible to have a better understanding on how the users perceive the recommended products [26]. Pappurajan, A. and S. P. Victor put forward an algorithm that judged whether the text in the Tweet is positive or negative based on the likelihood for each possibility to determine the sentiment of the text, whether it was positive or negative [27]. In order to facilitate sentiment analysis of user-generated content, Hogenboom et al. proposed to map sentiment conveyed by unstructured natural language text to universal star ratings, capturing intended sentiment. The results of our experiments revealed language-specific sentiment scores can separate universal classes of intended sentiment from one another to a

limited extent [28]. A new sentiment analysis technique called SentiPipe was presented, which took the best of a set of methods, generating a less sensitive analysis based context [29]. C. Clavel and Z. Callejas presented a comparative state of the art which analyzes the sentiment-related phenomena and the sentiment detection methods used in both communities and made an overview of the goals of socio-affective human-agent strategies [30].

3 Method

3.1 The basic principle of fuzzy clustering

Clustering analysis is to strictly divide each object into different classes. In fact, most of the objects are fuzzy. Here, the so-called fuzzy, mainly refers that the differences in objective things in the middle of the excessive nature is not clear, such as in the management system, medical diagnosis, artificial intelligence, image recognition, and other fields. Fuzzy set theory was first put forward by professor L. A. Zadeh; many meaningful results have been achieved until now. From the point of view of the development trend of the subject, it has extremely strong vitality and penetrability. Fuzzy set theory is applied to solving the problem of clustering in order to cluster fuzzy objects. The essence of fuzzy clustering is to construct fuzzy matrix according to the nature of the object itself, on the basis of this, according to a certain degree of membership, to determine the classification relationship.

The related definitions and basic principles of fuzzy clustering are as follows:

Definition 1: Let set of source information for statistical analysis be $X = (x_1, x_2, \dots, x_n)$, where x_i has m attributes, $(x_{i1}, x_{i2}, \dots, x_{im})$ represents a partition of x_i . X is $n \times m$ matrix, which is called initial value matrix.

Definition 2: In $X = (x_1, x_2, \dots, x_n)$, x_i and x_j ($i \neq j$) are any two different objects. The symbol r_{ij} stands for similarity degree between x_i and x_j ; r_{ij} is called as similarity coefficient.

Definition 3: Let U and V be two domains. For $(x, y) \in U \times V$, denote R as membership degree (or membership function) of U and V . $\mu_R(x, y) : U \times V \rightarrow [0, 1]$ represents fuzzy set R on U and V , which is a fuzzy relation from U to V .

Definition 4: Suppose U and V are finite domains, then all the r_{ij} , constitute fuzzy set R , which can be rewritten as $(r_{xy})_{n \times m}$, where $0 \leq r_{ij} \leq 1$ ($i \leq n, j \leq m$) and n and m are all positive integers. Matrix R is a fuzzy matrix and satisfies the following conditions:

- 1) Reflexivity: That is, any object must be the same as its own, which is denoted as $r_{ii} = 1$
- 2) Symmetry: If the object a is similar to the object b , then b should be similar to a , which is denoted as $r_{ij} = r_{ji}$

- 3) Transitivity: If the object a and object b are similar, object b and object c are similar, then a and c are similar, which is denoted as $R \circ R \subseteq R$

If R satisfies the above 1), 2), 3), then the relation represented by R is called as equivalence relation. So, fuzzy clustering analysis is carried out on the fuzzy equivalence relation.

$R = (r_{ij})_{n \times m}$ which satisfies the reflexivity, symmetry, and transitivity, is fuzzy equivalence matrix, also known as similarity matrix.

Definition 5: Let $R = (r_{ij})_{n \times m}$; for given $\lambda \in [0, 1]$, $R_\lambda = (r_{ij}(\lambda))_{n \times m}$ then R_λ is λ -cut matrix of R , where

$$r_{ij}(\lambda) = \begin{cases} 1, & r_{ij} \geq \lambda \\ 0, & r_{ij} < \lambda \end{cases}. \text{ The equivalent matrix } R = (r_{ij})_{n \times n}$$

can be used to classify the objects in domain, that is, when $R = (r_{ij})_{n \times n}$ is determined, for a given $\lambda \in [0, 1]$, the equivalent matrix R_λ can be obtained, and then get a level classification of λ .

3.2 Collection of micro-blog sentiment words set

In order to obtain the micro-blogging network sentiment words set before constructing special sentiment lexicon of micro-blog, the following steps are described:

- (1) Micro-blog corpus reprocessing

Before extracting the micro-blog network words, first of all, text information of micro-blog need to be reprocessed, such as micro-blog message: “//@ veggieg: flags can hang up a bit, dear,” “【# I'm a singer # live】zhang jie sings the great love of life and feeling...,” and so on. In the messages, “@user,” “# Theme Name#,” “punctuation.” and hyperlink information need to be removed in the reprocessing. n -gram ($n = 1, 2, 3, 4$) method is used to get micro-blog words set, then basic sentiment words set are filtered out; next, candidate networks' sentiment word set can be obtained. Finally, analyzing word frequency of the candidate word sets gotten by n -gram ($n = 2, 3, 4$), and setting a word frequency threshold value, effective candidate words set can be generated (if the candidate words frequency is lower than the threshold, it is filtered).

- (2) Discovery method of micro-blog network sentiment words (context-based entropy method)

Step 1: Compute each word in candidate word set with context-based entropy method. Let context word sets of a candidate word w be $L = \{l_1, l_2, l_3, \dots, l_p\}$ and $R = \{r_1, r_2, r_3, \dots, r_q\}$; its context entropy is computed as follows:

$$E_L(w) = -\frac{1}{n} \sum_{i=1}^p C(w, l_i) \log \frac{C(w, l_i)}{n}, l_i \in L \quad (1)$$

$$E_R(w) = -\frac{1}{m} \sum_{j=1}^q C(w, r_j) \log \frac{C(w, r_j)}{m}, r_j \in R \quad (2)$$

In formula 1 and 2, $n = \sum_{l_i \in L} C(w, l_i)$, $m = \sum_{r_j \in R} C(w, r_j)$,

where m and n represent the total frequency times of single word appearing before candidate word and after candidate word, respectively; $C(w, l_i)$ and $C(w, r_j)$ are respectively the frequency times of word; l_i and r_j appears in the candidate words w context.

Step 2: Determine network sentiment words. First, set context entropy threshold and count word frequency to extract users' network sentiment words in the micro-blog, then make up network sentiment words library of the micro-blog.

- (3) Filter of the network words (IF-IDF method)

Although some words have context entropy, they are not words like

“的是”, and “酱紫(这样子)”

which belongs to network word but has not the emotion color. For these disruptive words, they need to be filtered. In this paper, IF-IDF method is used to filter the network words. IF-IDF calculation method is shown in formula (3)

$$\text{TF-IDF}(w_i) = \text{freq}(w_i) \cdot \log \left(\frac{N}{df(w_i)} \right) \quad (3)$$

where

$\text{freq}(w_i)$ represents the occurrence times of word w_i in the corpus, N is total number of total documents in the corpus, and $df(w_i)$ is the frequency of documents containing w_i in the corpus.

In micro-blog corpus, the next steps for filtering network words are to calculate IF-IDF values of each candidate word, and then sort them, set a threshold value, and finally merge the candidate words higher than the threshold into micro-blog network sentiment word set.

3.3 A computation method for the sentiment value of sentiment phrase with modifier

In Chinese micro-blog message sentiment analysis, sentences are generally made up of sentiment phrases which contain sentiment words and its modifier. In this article, combining with the practical applications, the computation method of sentiment phrases trend generated by the following seven different combination modes is adopted (Table 1).

3.4 The sentiment intensity analysis of micro-blog content

Micro-blog content may be a phrase and may also be composed of two or more short sentences. To calculate sentiment intensity value of the whole micro-blog content, the first step is to analyze sentiment value of every phrase in micro-blog content; the second step is that if a complex sentence composed of two or more than two

Table 1 Computation method of sentiment phrases trend

Sequence number	Phrase combination mode	Formula of sentiment computing
1	$S = PW$	$E(PW)$
2	$S = NA + PW$	$a * E(PW) * E(NA)$, where a is adjustable parameter, its value range is from 1/2 to 1
3	$S = NA + NA + PW$	$E(PW) * E(NA) * E(NA)$
4	$S = DA + PW$	$(1 + E(DA)) * E(PW)$, $E(DA) < 0$ $E(PW) + (1 - E(PW)) * E(DA)$, $E(DA) > 0$ & $E(PW) > 0$ $E(PW) + (-1 - E(PW)) * E(DA)$, $E(DA) > 0$ & $E(PW) < 0$
5	$S = DA + DA + PW$	$E(PW) + (1 - E(PW)) * (E(DA1) + (1 - E(PW)) * (1 - E(PW)) * E(DA1)) * E(DA2)$
6	$S = NA + DA + PW$	$E(PW) + (1 - E(PW)) * (E(DA) - 0.2)$
7	$S = DA + NA + PW$	$E(PW) * E(NA) + (-1 - E(PW)) * E(NA) * E(DA)$

The meaning of each item in Table 1 is shown in Table 2

short sentences contains “coordinating,” “progressive,” “causal,” and “selected” relationship of conjunction, sentiment intensity value of the complex sentence is determined by means of accumulative sum of sentiment values; for the complex sentence with “hypothesis,” “conditions,” “concessions,” or “turning” relationship, sentiment intensity value of rear part of the sentence is taken as the sentiment intensity value of the whole sentence. Moreover, in a single short sentence, firstly, we should select sentiment words, and secondly, modify sentiment intensity value of sentiment words according to the weights of degree adverbs and negative words.

Define sentiment intensity value of sentiment word w as $E(w)$. Let sentiment intensity value of sentiment phrase with modifier be $ME(w)$. In order to identify modifier, we take w as the center, query before and after w until the nearest punctuation, and extract the modifier in the two adjacent punctuation area before and after w , then according to the rules of modification, take sentiment inclination correction analysis of a sentence. The algorithm is described as follows:

Algorithm name: sentiment intensity value computation based on sentence semantic with sentiment words

Input: micro-blog message sentence S

Output: sentiment intensity value $E(S)$ of message S

Step 1: Decompose the sentence into single Chinese words, then mark speech of words; finally, store the sequence of decomposed Chinese words

Step 2: Find the conjunction and punctuation in the sequence of decomposed words, obtain p short sequences $\{SS_k, k = 1, 2, \dots, p\}$, and store the conjunction types

Step 3: Compute sentiment intensity value of single phrase SS_k

Step 3.1: In sentiment library, search a sentiment word w and record its sentiment weight $E(w)$

Step 3.2: Extract the modifier in two adjacent punctuation area before and after sentiment word, get degree adverbs set $D = (D_1, D_2, \dots, D_n)$ and negative word sets $N = (N_1, N_2, \dots, N_m)$, then according to Table 2, obtain adjusted sentiment value of sentiment word w

with modifier, which is the sentiment intensity $ME(w)$ of phrase SS_k

Step 4: According to the phrase sequence relationship, use corresponding algorithm to obtain sentiment intensity values $E(S)$ of the whole sentence

Step 5: Determine sentence types

Case 1: question sentence, then $E(S) = -E(S)$, turn step 6

Case 2: exclamatory sentences, then $E(S) = 2E(S)$, turn step 6

Step 6: output $E(S)$

3.5 Generation of sentiment dynamic clustering graph

Fuzzy clustering concept is based on general clustering. Assume that the set of X is classified to many classes by membership function; each class is thought as a fuzzy subset of X ; the classification matrix corresponding to each classification is a fuzzy matrix R .

Cluster analysis is to use the similarity scale to measure the closeness degree between things and to realize classification. The essence of cluster analysis is to construct fuzzy matrix based on the attribute of research object itself and, on this basis, determine classification relationship according to certain membership. The main steps include determining sample statistic index (feature extraction, establish the original data matrix), data standardization (normalized), and distance calibration to establish fuzzy similar matrix and computing fuzzy equivalence matrix; the detailed steps are as follows:

Table 2 Meaning of emotional phrase tag

Tag	Sentiment formula
PW	Sentiment words
NA	Negative words
DA	Adverbs of degree
$E(*)$	*Sentiment intensity value of the word, *stands for negative emotional words and adverbs of degree
S	The phrases made up of modifier and sentiment words

(1) Construction of original data matrix

Let classified objects be domain $X = (x_1, x_2, \dots, x_n)$ and n represents the number of objects; each object has m attribute features, $x_i = (x_{i1}, x_{i2}, \dots, x_{im})$, $i = 1, 2, \dots, n$. So, original data matrix is given as follows:

$$A = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1m} \\ x_{21} & x_{22} & \dots & x_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \dots & x_{nm} \end{bmatrix}$$

where x_{nm} represents original data of the m th feature of the n th classified object.

(2) Standardization processing of data

Because of the different orders of magnitude and dimension of feature indexes, for the convenience of analysis and comparison, the sample data needs to be standardized; data range is transferred to the domain $[0, 1]$ (normalized processing). Assume that there are n sample objects x_1, x_2, \dots, x_n , each object has m feature values y_1, y_2, \dots, y_m ; x_{ij} means the j th attribute value of the i th object. The following moving range formula (4) is used in standardization method:

$$x'_{ij} = \frac{x_{ij} - x_{\min}}{x_{\max} - x_{\min}} \quad (4)$$

(3) Construction of a fuzzy similar matrix

Let the domain set be $X = (x_1, x_2, \dots, x_n)$, $x_i = (x_{i1}, x_{i2}, \dots, x_{im})$, in accordance with the commonly clustering method; we can determine fuzzy similarity coefficient r_{ij} , which is similar degree of x_i, x_j ($r_{ij} = R(x_i, x_j)$).

The following 12 methods can be used to determine r_{ij} , which can be chosen for actual needs.

(1) Maximum minimum method

$$r_{ij} = \frac{\sum_{k=1}^m (x_{ik} \wedge x_{jk})}{\sum_{k=1}^m (x_{ik} \vee x_{jk})} \quad (5)$$

(2) Arithmetic average minimum method

$$r_{ij} = \frac{\sum_{k=1}^m (x_{ik} \wedge x_{jk})}{\frac{1}{2} \sum_{k=1}^m (x_{ik} + x_{jk})} \quad (6)$$

(3) Geometric average minimum method

$$r_{ij} = \frac{\sum_{k=1}^m (x_{ik} \wedge x_{jk})}{\sum_{k=1}^m \sqrt{x_{ik} x_{jk}}} \quad (7)$$

(4) Index similarity coefficient method

$$r_{ij} = \frac{1}{m} \sum_{k=1}^m e^{-\frac{4(x_{ik} - x_{jk})^2}{3s_k^2}} \quad (8)$$

where $s_k = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_{ik} - \bar{x}_k)^2}$, $\bar{x}_k = \frac{1}{n} \sum_{i=1}^n x_{ik}$

(5) Correlation coefficient method

$$r_{ij} = \frac{\sum_{k=1}^m |x_{ik} - \bar{x}_i| |x_{jk} - \bar{x}_j|}{\sqrt{\sum_{k=1}^m (x_{ik} - \bar{x}_i)^2} \cdot \sqrt{\sum_{k=1}^m (x_{jk} - \bar{x}_j)^2}} \quad (9)$$

Where $\bar{x}_i = \frac{1}{m} \sum_{k=1}^m x_{ik}$, $\bar{x}_j = \frac{1}{m} \sum_{k=1}^m x_{jk}$

(6) Chebyshev distance method

$$r_{ij} = 1 - C(d(x_i, x_j))^a \quad (10)$$

where $d(x_i, x_j) = \max_{k=1}^m |x_{ik} - x_{jk}|$, $a = 1$

(7) Absolute value index method

$$r_{ij} = e^{-\sum_{k=1}^m |x_{ik} - x_{jk}|} \quad (11)$$

(8) Absolute value inverse method

$$r_{ij} = \begin{cases} 1 & i = j \\ \frac{C}{\sum_{k=1}^m |x_{ik} - x_{jk}|} & i \neq j \end{cases} \quad (12)$$

where C is proper positive number, satisfy

$$C \leq \min_{i,j} \left(\sum_{k=1}^m |x_{ik} - x_{jk}| \right)$$

(9) Dot product method

$$r_{ij} = \begin{cases} 1 & i = j \\ \frac{1}{C} \sum_{k=1}^m x_{ik} x_{jk} & i \neq j \end{cases} \quad (13)$$

where C is proper positive number, satisfy

$$C \geq \max_{i,j} \left(\sum_{k=1}^m x_{ik} x_{jk} \right)$$

(10) Absolute value reduction method

$$r_{ij} = \begin{cases} 1 & i = j \\ 1 - C \sum_{k=1}^m |x_{ik} - x_{jk}| & i \neq j \end{cases} \quad (14)$$

where C is a proper positive number, which satisfies $0 \leq r_{ij} \leq 1$.

(11) Angle cosine method

$$r_{ij} = \frac{\sum_{k=1}^m x_{ik} x_{jk}}{\sqrt{\sum_{k=1}^m x_{ik}^2} \cdot \sqrt{\sum_{k=1}^m x_{jk}^2}} \quad (15)$$

(12) Subjective evaluation method

Experts with rich experience experts directly assess the similarity degree between x_i and x_j and give the value of r_{ij} ($0 \leq r_{ij} \leq 1$).

(4) Clustering (fuzzy equivalence matrix)

The fuzzy similar matrix constructed in the above section usually has reflexivity and symmetry and does not satisfy transitivity. In order to classify, we need to obtain fuzzy equivalence matrix by modifying fuzzy similar matrix. The commonly used method is the transitive closure method, that is $R \rightarrow R^2 \rightarrow R^4 \rightarrow \dots \rightarrow R^{2^k}$ until $R^{2^k} = (R^{2^k})^2$, then R^{2^k} is called transitive closure. After obtaining the transitive closure, we can construct fuzzy equivalence matrix in domain set X . Then, set $\lambda \in [0, 1]$, find R_λ , namely, under different levels of λ , a dynamic clustering graph in range X can be generated by data processing tool such as Matlab, SPSS, and so on.

4 Sentiment analysis for QQ zone users

4.1 Original data matrix

Each row in above matrix represents a QQ zone users; each column represents sentiment values in a certain period. By using the Chinese text sentiment computation method, we can calculate sentiment values of 15 QQ zone users in 10 different periods. So the original data matrix is transformed into sentiment matrix of QQ zone; the matrix elements are as shown in Table 3.

4.2 Sample data standardization processing

Sample data standardization is normalization of data; the data are compressed in the range $[0, 1]$. By using moving range formula (formula 4), standardization matrix are obtained by Matlab programming, the matrix elements are shown in Table 4.

4.3 Construction of fuzzy similar matrix

Here, absolute value reduction method is adopted to determine the similarity degree between two samples, namely, the similarity coefficient r_{ij} (see formula (14)). C is set to 0.1, and then 15×15 sentiment fuzzy similar matrix of 15 QQ zone users can be obtained; the matrix elements are shown in Table 5.

4.4 Sentiment clustering

The fuzzy similar matrix got from the above section does not necessarily have transitivity; before clustering, transitive closure $t(R)$ needs to be found by least squares method, namely fuzzy equivalence matrix R' is obtained. The matrix elements are shown in Table 6. Then, let the value of λ be changed from high to low; dynamic clustering graph can be generated in Fig. 1.

4.5 Clustering results analysis

It can be seen from Fig. 1:

- (1) When λ is set to 0.9257, emotions of 15 users are classified into 14 classes: the 11th and 12th users are the same class. Table 1 shows that the 11th and 12th rows in the original data matrix are all positive; this reveals that the two users have positive emotion at the time of the study, and other Internet users are always in the negative emotion state in some time, namely the negative emotion. It also can be seen that the objects which get together for a class first have the higher degree of similarity and closer emotions tend. Analysis shows that the fuzzy clustering algorithm applied in sentiment analysis has a good effect.

Table 3 Original data matrix

0.5	0.55	0.69	0.58	0.57	0.45	-0.14	-0.89	0.45	0.7
0.5	0.54	0.69	0.69	0.57	0.95	1	-0.69	0.85	0.9
0.89	0.78	0.89	0.69	0.57	-0.95	-0.89	-0.69	0.75	0.9
0.1	-0.78	-0.9	-0.69	0.57	-1	-0.9	-0.69	0.23	0.3
-0.1	-0.2	-0.3	-0.59	0.86	1	0.9	-0.69	0.23	-0.3
0.1	-0.2	0.3	0.52	0.66	0.8	0.9	0.69	0.73	-0.36
-0.1	0.2	0.3	0.52	0.46	-0.8	0.1	-0.69	0.73	-0.36
0.5	0.55	0.89	0.2	-0.3	0.45	-0.14	-0.89	0.45	0.7
0.19	0.85	0.79	0.2	-0.3	-0.45	-0.14	-0.89	0.45	0.7
0.19	0.25	0.74	0.62	-0.33	-0.45	-0.14	-0.59	-0.45	0.7
0.9	0.5	0.74	0.62	0.33	0.45	0.14	0.59	0.45	0.39
0.9	0.89	0.74	0.62	0.33	0.89	0.55	0.59	0.45	0.29
0.9	0.89	0.84	0.92	-0.13	0.89	0.85	0.78	-0.85	0.3
0.19	0.89	0.44	0.82	-0.33	0.59	0.65	0.78	-0.45	0.13
1	0.89	0.39	0.99	0.85	1	1	1	0.82	1

In above table, the emotional intensity values are controlled between $[-1, 1]$; when the values are greater than zero, it is regarded as the positive feelings; if the values are less than zero, it is the negative emotions

Table 4 Data standardization matrix

0.545455	0.796407	0.888268	0.755952	0.756303	0.725	0.4	0	0.764706	0.779412
0.545455	0.790419	0.888268	0.821429	0.756303	0.975	1	0.10582	1	0.926471
0.9	0.934132	1	0.821429	0.756303	0.025	0.005263	0.10582	0.941176	0.926471
0.181818	0	0	0	0.756303	0	0	0.10582	0.635294	0.485294
0	0.347305	0.335196	0.059524	1	1	0.947368	0.10582	0.635294	0.044118
0.181818	0.347305	0.670391	0.720238	0.831933	0.9	0.947368	0.835979	0.929412	0
0	0.586826	0.670391	0.720238	0.663866	0.1	0.526316	0.10582	0.929412	0
0.545455	0.796407	1	0.529762	0.02521	0.725	0.4	0	0.764706	0.779412
0.263636	0.976048	0.944134	0.529762	0.02521	0.275	0.4	0	0.764706	0.779412
0.263636	0.616766	0.916201	0.779762	0	0.275	0.4	0.15873	0.235294	0.779412
0.909091	0.766467	0.916201	0.779762	0.554622	0.725	0.547368	0.783069	0.764706	0.551471
0.909091	1	0.916201	0.779762	0.554622	0.945	0.763158	0.783069	0.764706	0.477941
0.909091	1	0.972067	0.958333	0.168067	0.945	0.921053	0.883598	0	0.485294
0.263636	1	0.748603	0.89881	0	0.795	0.815789	0.883598	0.235294	0.360294
1	1	0.72067	1	0.991597	1	1	1	0.982353	1

(2) Users can select the proper threshold to search for friends that are similar or contrary to their emotions, according to their circumstance and achieve the emotional pursuit of life. This illustrates that the dynamic fuzzy clustering analysis method has good maneuverability and flexibility in the analysis of emotion.

5 Sentiment clustering analysis of micro-blog users on fuzzy clustering algorithm

5.1 Data preparation

Experimental data were from micro-blog of www.sina.com. A large number of users micro-blog messages with user ID were collected by crawler system, and the time, website,

and forwarding times of the messages were stored in a database. Then, the messages were reprocessed and computed by sentiment algorithm; we can obtain the sentiment values of the messages in different periods. In the experiment, we selected 50 micro-blog users and found their sentiment values in 10 periods. Denote object domain as $X = \{x_1, x_2, \dots, x_n\}$; let object attributes of x_i be $(x_{i1}, x_{i2}, \dots, x_{im})$, where $n = 50$, $m = 10$.

5.2 Fuzzy clustering

Based on fuzzy clustering theory knowledge, we make the original data normalization and construct fuzzy equivalence matrix; finally, generated dynamic fuzzy clustering diagram show in Fig. 2.

Table 5 Fuzzy similar matrix

1	0.859	0.781	0.554	0.572	0.636	0.71	0.893	0.808	0.759	0.819	0.751	0.587	0.616	0.638
0.859	1	0.739	0.436	0.613	0.694	0.649	0.752	0.667	0.645	0.735	0.753	0.629	0.606	0.774
0.781	0.739	1	0.575	0.374	0.47	0.67	0.696	0.717	0.661	0.703	0.662	0.518	0.458	0.616
0.554	0.436	0.575	1	0.644	0.483	0.634	0.492	0.553	0.543	0.446	0.385	0.247	0.36	0.247
0.572	0.613	0.374	0.644	1	0.749	0.677	0.51	0.481	0.471	0.494	0.521	0.414	0.5	0.476
0.636	0.694	0.47	0.483	0.749	1	0.746	0.536	0.507	0.519	0.693	0.711	0.588	0.684	0.667
0.71	0.649	0.67	0.634	0.677	0.746	1	0.628	0.689	0.691	0.646	0.586	0.421	0.559	0.461
0.893	0.752	0.696	0.492	0.51	0.536	0.628	1	0.903	0.804	0.758	0.69	0.615	0.655	0.531
0.808	0.667	0.717	0.553	0.481	0.507	0.689	0.903	1	0.865	0.673	0.641	0.56	0.662	0.482
0.759	0.645	0.661	0.543	0.471	0.519	0.691	0.804	0.865	1	0.667	0.593	0.612	0.725	0.434
0.819	0.735	0.703	0.446	0.494	0.693	0.646	0.758	0.673	0.667	1	0.926	0.762	0.712	0.721
0.751	0.753	0.662	0.385	0.521	0.711	0.586	0.69	0.641	0.593	0.926	1	0.835	0.756	0.781
0.587	0.629	0.518	0.247	0.414	0.588	0.421	0.615	0.56	0.612	0.762	0.835	1	0.829	0.705
0.616	0.606	0.458	0.36	0.5	0.684	0.559	0.655	0.662	0.725	0.712	0.756	0.829	1	0.625
0.638	0.774	0.616	0.247	0.476	0.667	0.461	0.531	0.482	0.434	0.721	0.781	0.705	0.625	1

Table 6 Fuzzy equivalence matrix

1	0.859	0.7806	0.6343	0.6769	0.7113	0.7098	0.8931	0.8931	0.8075	0.8195	0.8195	0.762	0.7512	0.7744
0.859	1	0.7806	0.6343	0.6944	0.7113	0.7098	0.859	0.8075	0.7592	0.8195	0.7744	0.7528	0.7528	0.7744
0.7806	0.7806	1	0.6343	0.6695	0.6944	0.7098	0.7806	0.7806	0.7592	0.7806	0.7512	0.7025	0.7025	0.7386
0.6343	0.6343	0.6343	1	0.6444	0.6444	0.6444	0.6285	0.6343	0.6343	0.6343	0.5861	0.5541	0.5586	0.5749
0.6769	0.6944	0.6695	0.6444	1	0.7486	0.7459	0.6285	0.6769	0.6769	0.6927	0.7113	0.6128	0.6838	0.667
0.7113	0.7113	0.6944	0.6444	0.7486	1	0.7459	0.6944	0.6893	0.6909	0.7113	0.7113	0.7113	0.7113	0.7113
0.7098	0.7098	0.7098	0.6444	0.7459	0.7459	1	0.7098	0.7098	0.7098	0.7098	0.7113	0.6457	0.6909	0.667
0.8931	0.859	0.7806	0.6285	0.6285	0.6944	0.7098	1	0.9033	0.8649	0.8195	0.7585	0.7585	0.725	0.7521
0.8931	0.8075	0.7806	0.6343	0.6769	0.6893	0.7098	0.9033	1	0.8649	0.8075	0.7512	0.6729	0.725	0.6729
0.8075	0.7592	0.7592	0.6343	0.6769	0.6909	0.7098	0.8649	0.8649	1	0.7592	0.7512	0.725	0.725	0.6671
0.8195	0.8195	0.7806	0.6343	0.6927	0.7113	0.7098	0.8195	0.8075	0.7592	1	0.9257	0.8349	0.762	0.7808
0.8195	0.7744	0.7512	0.5861	0.7113	0.7113	0.7113	0.7585	0.7512	0.7512	0.9257	1	0.8349	0.8288	0.7808
0.762	0.7528	0.7025	0.5541	0.6128	0.7113	0.6457	0.7585	0.6729	0.725	0.8349	0.8349	1	0.8288	0.7808
0.7512	0.7528	0.7025	0.5586	0.6838	0.7113	0.6909	0.725	0.725	0.725	0.762	0.8288	0.8288	1	0.7563
0.7744	0.7744	0.7386	0.5749	0.667	0.7113	0.667	0.7521	0.6729	0.6671	0.7808	0.7808	0.7808	0.7563	1

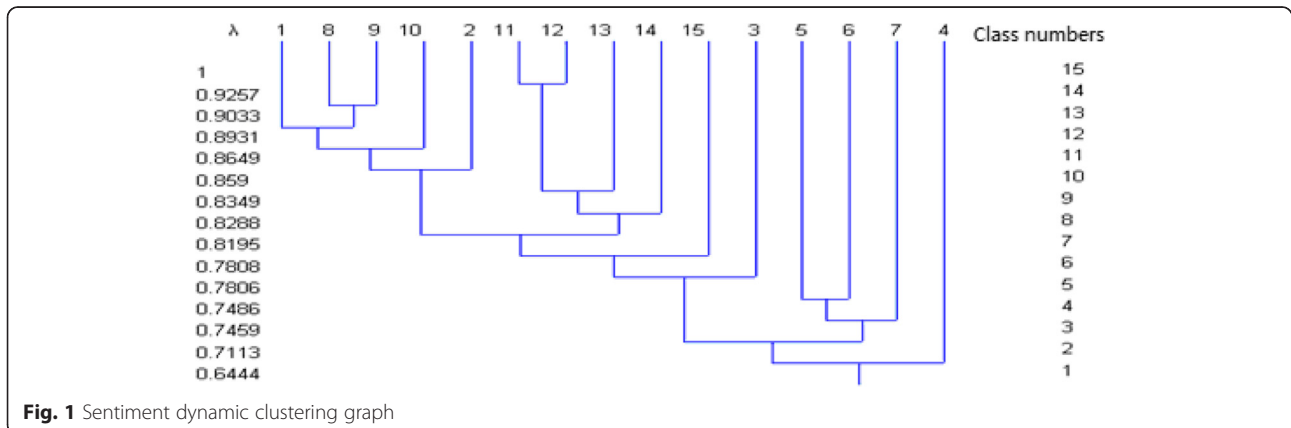
In Fig. 2, users' sentiment change is illustrated by calculating the user network news; no matter what social circles users are in, they will show their emotions, so sentiment analysis is not directly related to social categories. As illustrated in Fig. 2, when clustering threshold λ is more close to 1, classification number is more, λ is small to a certain value, and the samples will belong to one class (λ is set to 0.72; all users are incorporated into one category). The advantage of the clustering method is that λ can be selected according to the actual needs in order to get the appropriate classification.

Denote classification number as r , as shown in Fig. 2, the classification number r is respectively 1, 2, 6, 7, 14, 16, 19, 22, 25, 31, 35, 37, 38, 40, 43, 44, 45, 46, 47, 48, 50. $r = 1$ represents that all 50 users is in one class; $r = 50$ represents that each user is in one class respectively, namely 50 classes. Five kinds of classification are analyzed as follows:

- 1) $r = 2: \{x_{21}\}, \{x_1-x_{20}, x_{22}-x_{50}\}$.
- 2) $r = 6: \{x_{21}\}, \{x_{34}\}, \{x_{29}, x_{49}\}, \{x_6\}, \{x_4\}, \{x_1-x_3, x_5, x_7-x_{20}, x_{22}-x_{28}, x_{30}-x_{33}, x_{35}-x_{48}, x_{50}\}$.
- 3) $r = 7: \{x_{21}\}, \{x_{34}\}, \{x_{29}, x_{49}\}, \{x_6\}, \{x_4\}, \{x_{47}\}, \{x_1-x_3, x_5, x_7-x_{20}, x_{22}-x_{28}, x_{30}-x_{33}, x_{35}-x_{48}, x_{50}\}$.
- 4) $r = 14: \{x_{21}\}, \{x_{34}\}, \{x_{29}, x_{49}\}, \{x_6\}, \{x_4\}, \{x_{47}\}, \{x_{50}\}, \{x_{38}\}, \{x_{23}, x_{43}\}, \{x_{14}, x_{41}\}, \{x_8\}, \{x_7\}, \{x_3\}, \{x_1, x_2, x_5, x_9-x_{13}, x_{15}-x_{20}, x_{22}, x_{24}-x_{28}, x_{30}-x_{33}, x_{35}-x_{37}, x_{39}-x_{40}, x_{42}, x_{44}-x_{46}, x_{48}\}$.
- 5) $r = 16: \{x_{21}\}, \{x_{34}\}, \{x_{29}, x_{49}\}, \{x_6\}, \{x_4\}, \{x_{47}\}, \{x_{50}\}, \{x_{38}\}, \{x_{23}, x_{43}\}, \{x_{14}, x_{41}\}, \{x_8\}, \{x_7\}, \{x_3\}, \{x_{12}, x_{27}\}, \{x_9\}, \{x_1, x_2, x_5, x_{10}-x_{11}, x_{13}, x_{15}-x_{20}, x_{22}, x_{24}-x_{26}, x_{28}, x_{30}-x_{33}, x_{35}-x_{37}, x_{39}, x_{40}, x_{42}, x_{44}-x_{46}, x_{48}\}$.

5.3 Clustering effect test and result analysis

In view of the above five kinds of classification results, in order to determine the optimum classification, F test is used to F statistics: define the classification number as r , sample number of the no. j class is n_j , the sample of

**Fig. 1** Sentiment dynamic clustering graph

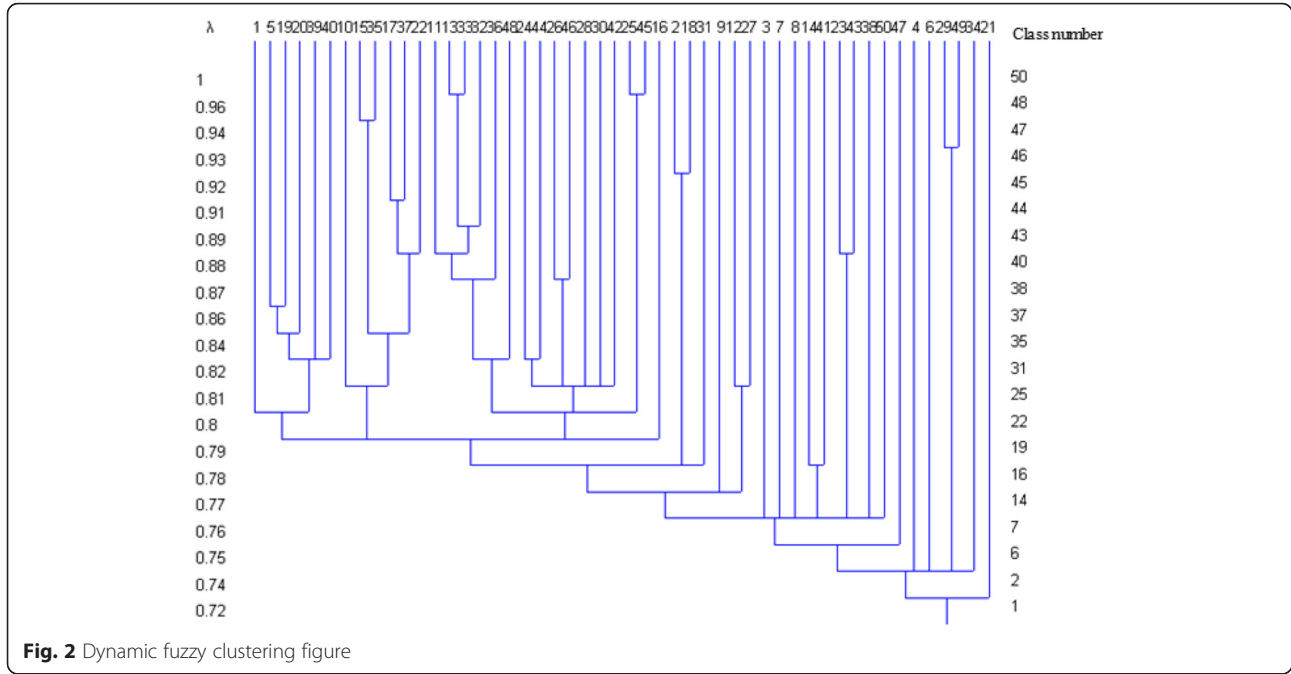


Fig. 2 Dynamic fuzzy clustering figure

the j th class is denoted as $x_1^{(j)}, x_2^{(j)}, \dots, x_{n_j}^{(j)}$; clustering center, the j th class is a vector $\bar{x}^{(j)} = (\bar{x}_1^{(j)}, \bar{x}_2^{(j)}, \dots, \bar{x}_m^{(j)})$, which $\bar{x}_k^{(j)}$ is the average of the k th feature, namely

$$\bar{x}_k^{(j)} = \frac{1}{n_j} \sum_{i=1}^{n_j} x_{ik}^{(j)}, \quad (k = 1, 2, \dots, m) \quad (16)$$

$$F = \frac{\sum_{j=1}^r n_j \|\bar{x}^{(j)} - \bar{x}\| / (r-1)}{\sum_{j=1}^r \sum_{i=1}^{n_j} \|x_i^{(j)} - \bar{x}^{(j)}\| / (n-r)} \quad (17)$$

where $\|\bar{x}^{(j)} - \bar{x}\| = \sqrt{\sum_{k=1}^m (\bar{x}_k^{(j)} - \bar{x}_k)^2}$ stands for the distance between $\bar{x}^{(j)}$ and \bar{x} ; $\|x_i^{(j)} - \bar{x}^{(j)}\|$ represents the distance between the i th sample in the j th class, $x_i^{(j)}$ and its center, $\bar{x}^{(j)}$, the F value obeys the F distribution in which the pair of degrees of freedom is $r-1$, $n-r$, the numerator of F stands for the distance between two classes, and the denominator of F represents the distance between each sample in the class. The greater the F value is, the greater the distance between two classes is, namely, the greater the difference between two classes is, the better the classification results are.

Based on the above principle, we can obtain the following results:

- 1) If $r = 2$, then $n_1 = 1$, $n_2 = 49$, we can obtain $F_2 = 58.457\%$
- 2) If $r = 6$, then $n_1 = 1$, $n_2 = 1$, $n_3 = 2$, $n_4 = 1$, $n_5 = 1$, $n_6 = 44$, $F_6 = 76.111\%$ can be computed.

- 3) If $r = 14$, then $n_1 = 1$, $n_2 = 1$, $n_3 = 2$, $n_4 = 1$, $n_5 = 1$, $n_6 = 1$, $n_7 = 1$, $n_8 = 1$, $n_9 = 2$, $n_{10} = 2$, $n_{11} = 1$, $n_{12} = 1$, $n_{13} = 1$, $n_{14} = 33$, we can derive $F_{14} = 58.441\%$.
- 4) If $r = 16$, $n_1 = 1$, $n_2 = 1$, $n_3 = 2$, $n_4 = 1$, $n_5 = 1$, $n_6 = 1$, $n_7 = 1$, $n_8 = 1$, $n_9 = 2$, $n_{10} = 2$, $n_{11} = 1$, $n_{12} = 1$, $n_{13} = 1$, $n_{14} = 2$, $n_{15} = 1$, $n_{16} = 30$, $F_{16} = 61.236\%$ can be obtained.

From above data, we can get $F_6 > F_{16} > F_2 > F_{14}$, so dividing into six categories is the best classification method.

When 50 users are divided into six classes, sentiment tendency graph of each user is shown in Fig. 2.

Figure 3 illustrates that emotional states of the users in the first class are not very good as a whole; emotional value is below zero value in most cases, namely negative emotions are more serious. Therefore, relevant departments should take corresponding measures, to avoid users do something radical actions and to maintain social harmony and stability. Moods of users in the second category is a bit low early in the month, then, is slowly and slightly better; although there are some fluctuations of moods, the user's emotional values are overall positive; the users belong to the normal situation. Two users' emotional trends in the third class are consistent, except sentiment values were close to -1 at 2014/12/04; their negative affection degrees were strong, and their emotions are preferred in most times. And so on, we can analyze the user's emotional state in the fourth, fifth, and sixth class. Through the emotional charts, we can not only understand the emotional change

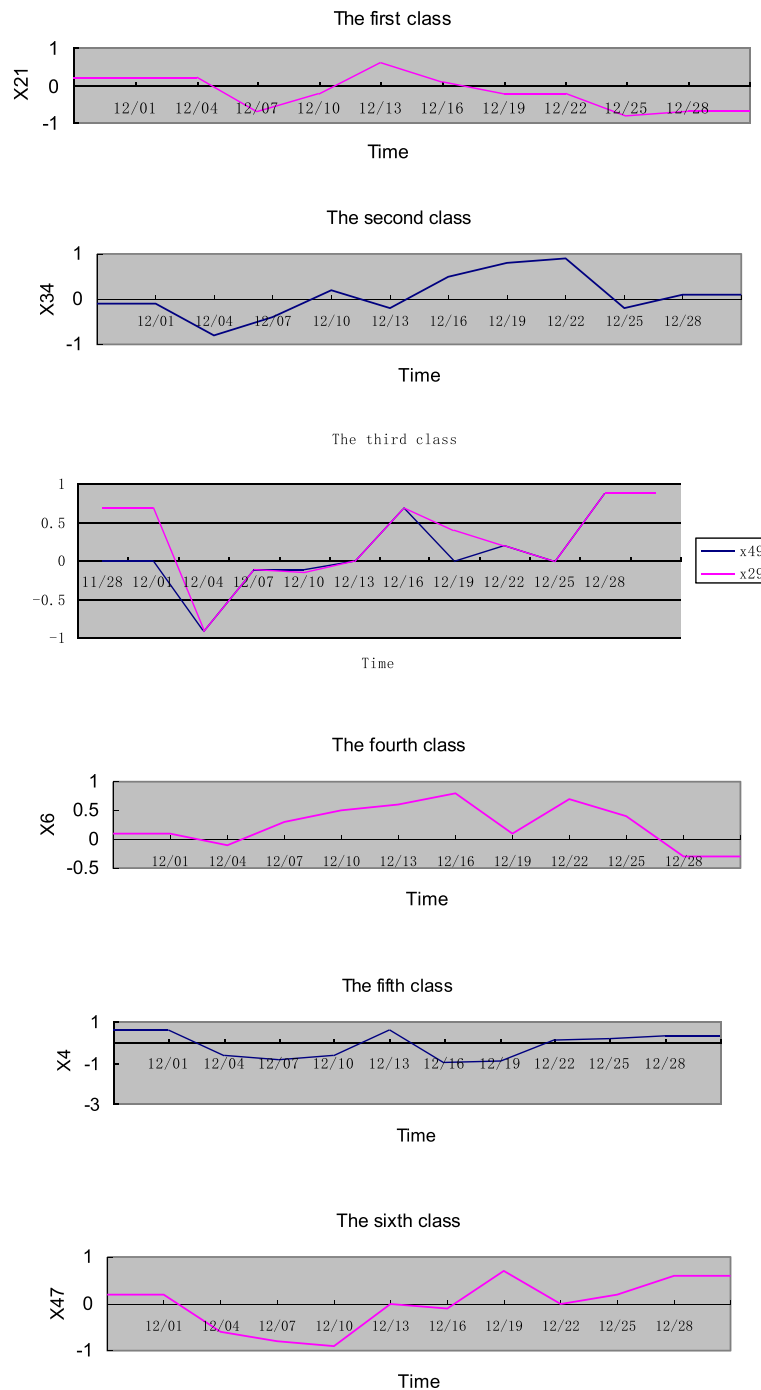


Fig. 3 Users sentiment trend

of individual users but also find users within the similar emotional states by different classifications.

6 Application examples

In recent years, fuzzy clustering analysis is widely used in data mining, pattern recognition, and comprehensive index evaluation. In this paper, the fuzzy clustering

algorithm is applied to sentiment analysis of Sina microblog and Tencent QQ space users, which achieves good results.

- (1) Sentiment analysis of Tencent QQ space users
 - (a) Collect the news sent by 15 QQ users in 10 different time periods

(b) For the selected time period, the period of time is adjustable; for highly active users, a short time interval can be selected; on the contrary, for low active users, a longer time interval can be selected, and if there exist a number of news in each time, take the average of the news as the value of news. If there is no published news, let the value of news be 0

(c) Calculate emotional value of 15 QQ users in 10 different time periods with emotional Chinese text calculation method mentioned above to construct a dynamic fuzzy clustering map

The dynamic fuzzy clustering map shows some samples relatively belonging to a certain class in some extent, which can be used to make appropriate decisions according to their own preferences by users and to forecast users' emotional trend with the curve fitting method and take appropriate preventive measures for some emotional extreme or abnormal people, to promote social harmony, stability, and development.

(2) Sentiment analysis of micro-blog Sina user

There are three steps as follows:

(a) Design of micro-blog Sina data acquisition system
After investigating and analyzing API of micro-blog Sina open platform, we obtain HTML structure of the micro-blog and parse micro-blog news. Next, using web crawler technology, according to news features of micro-blog Sina, web crawler data acquisition system for micro-blog Sina is designed and implemented.

(b) The construction process of micro-blog the emotion lexicon
We calculate emotional intensity (weights) of the basic emotional words and construct a basic emotion lexicon with Hownet-based word similarity computation method; finally, the network emotion lexicon and micro-blog emotion lexicon are obtained.

(c) Classification analysis of micro-blog user emotions
The numerical value of emotional information is defined as the range of $[-1, 1]$, and when the emotional value is greater than 0, the user's emotion is taken as the positive emotion, on the contrary, the emotional value which is less than 0 was for the negative emotions. Then, the fuzzy clustering algorithm is used to cluster the emotional values, and emotional value of the whole blog information is calculated based on the emotional words with modifiers and other factors. Next, emotional values of 50 users in 10 time periods are selected to make emotion classification analysis for micro-blog users by using the fuzzy clustering algorithm, and F testing method is used

to calculate an optimal classification. Finally, we predict the classification results with SPSS tool and generate emotion chart to understand more intuitively the user's emotional state.

7 Conclusions

Sentiment analysis has become the inevitable trend of big data development. Quantizing the human's emotion to enable computer to identify and analysis automatically has become one of the most important hot research topics of many scholars. In the paper, fuzzy clustering algorithm is applied to analyze emotional change of Sina micro-blog users. Firstly, sentiment intensity of text information issued by micro-blog users is analyzed. Next, combined with sentiment words and influence of modifiers on sentiment intensity, sentiment value of the whole text is calculated. Finally, Fuzzy clustering algorithm is used to analyze sentiment classification of micro-blog users, and classification result is forecasted with tools such as SPSS, and sentiment trend charts are generated to understand users' emotion state more intuitively. The research provides an important basis for decision-making to relevant research departments.

Competing interests

The authors declare that they have no competing interests.

Acknowledgements

This work was supported by National Natural Science Foundation Project under grants 61175122, as well as by the Applied Basic Research Project of Sichuan province of China (2013JY0134) and by Key Project of Sichuan Educational Commission (No. 15ZA0049).

Received: 27 January 2016 Accepted: 24 April 2016

Published online: 06 May 2016

References

1. F Liu, W Xian, Micro-blog builds a new platform for mobile learning. *China Educ. Technol. Equip.* 26 (36), 26–27 (2012).
2. X Song, SW features analysis in micro-blog. *J. CHIFENG Univ.* 29 (1), 96–98 (2013).
3. H Wang et al., Literature review of sentiment classification on Web text. *J. Chin. Soc. Sci. Tech. Inf.* 29 (5), 931–938 (2010).
4. Y Zhao, B Qin, T Liu, Sentiment analysis on text. *J. Softw.* 21(8), 1834–1848 (2010).
5. A Go, R Bhayani, L Huang, *Twitter sentiment classification using distant supervision*. *Cs224n Project Report*, 2009, pp. 1–12
6. B O'Connor, R Balasubramanyam, BR Routledge, NA Smith, *From Tweets to Polls: Linking Text Sentiment to Public Opinion Time Series*, Conference: *Proceedings of the Fourth International Conference on Weblogs and Social Media*, 2010, pp. 122–129
7. A Pak, P Paroubek, *Twitter as a Corpus for Sentiment Analysis and Opinion Mining* [C]/Seventh Conference on International Language Resources and Evaluation, 2010
8. A Bifet, E Frank, Sentiment knowledge discovery in Twitter streaming data. *Lect. Notes Comput. Sci.* 6332, 1–15 (2010)
9. D Davidov, O Tsur, A Rappoport, *Enhanced Sentiment Learning Using Twitter Hashtags and Smileys*. *International Conference on Computational Linguistics: Posters*, 2010, pp. 241–249
10. CX Lin et al., *PET: A Statistical Model for Popular Events Tracking in Social Communities*. *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2010, pp. 929–938

11. A Celikyilmaz, D Hakkani-Tur, J Feng, *Probabilistic Model-Based Sentiment Analysis of Twitter Messages*. *Spoken Language Technology Workshop (SLT), 2010 IEEE IEEE*, 2010, pp. 79–84
12. L Barbosa, J Feng, *Robust Sentiment Detection on Twitter from Biased and Noisy Data*. *23rd International Conference on Computational Linguistics*, vol. 23, 2010, pp. 36–44
13. J Bollen, A Pepe, H Mao, Modeling public mood and emotion: Twitter sentiment and socio-economic phenomena. *Biochem. Pharmacol.* **44**(12), 2365–2370 (2010)
14. L Jiang, M Yu, M Zhou et al., *Target-Dependent Twitter Sentiment Classification: Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, vol. 1, 2011, pp. 151–160
15. L Zhang et al., *Combining Lexicon-Based and Learning-based Methods for Twitter Sentiment Analysis*. *HP Laboratories Technical Report*, 2011
16. M E Hoque, D J McDuff, R W Picard, Exploring temporal patterns in classifying frustrated and delighted smiles. *IEEE Trans Affect Comput.* **3**(3), 323–334 (2012).
17. DJ McDuff, R el Kaliouby, RW Picard, Crowdsourcing facial responses to online videos. *IEEE Trans. Affect. Comput.* **3**(4), 456–468 (2012)
18. DN Trung et al., "Towards Modeling Fuzzy Propagation for Sentiment Analysis in Online Social Networks: A Case Study on TweetScope." *Cognitive Infocommunications (CogInfoCom)*, 2013 *IEEE 4th International Conference on IEEE*, 2013, pp. 331–338
19. S Venkatesh, D Phung, D Bo, M Berk, Affective and content analysis of online depression communities. *IEEE Trans. Affect. Comput.* **5**(3), 217–226 (2014)
20. TH Nguyen, K Shiraia, J Velcinb, Sentiment analysis on social media for stock movement prediction. *Expert Syst. Appl.* **42**, 9603–9611 (2015)
21. DN Trung, JJ Jung, Sentiment analysis based on fuzzy propagation in online social networks: a case study on TweetScope. *Comput Sci. Inf. Syst.* **11**(1), 215–228 (2014)
22. W Xing, Z Shaojian, Chinese text sentiment analysis utilizing emotion degree lexicon and fuzzy semantic model. *Int. J. Softw. Sci. Comput. Intell.* **6**(4), 20–32 (2014)
23. Z Wang et al., *Issues of Social Data Analytics with a New Method for Sentiment Analysis of Social Media Data*, *Cloud Computing Technology and Science (CloudCom)*, 2014 *IEEE 6th International Conference on IEEE*, 2014, pp. 899–904
24. MD Munezero, CS Montero, E Sutinen, J Pajunen, Are they different? Affect, feeling, emotion, sentiment, and opinion detection in text. *IEEE Trans. Affect. Comput.* **5**(2), 101–111 (2014)
25. H I Ahn, R W Picard, Measuring affective-cognitive experience and predicting market success. *IEEE Trans. Affect. Comput.* **5**(2):173–186 (2014).
26. F Santos-Sanchez, A Mendez-Vazquez, "Sentiment Analysis for e-Services." *Advanced Applied Informatics (IIAIAI)*, 2014 *IIAI 3rd International Conference on IEEE*, 2014, pp. 42–47
27. A Pappurajan, SP Victor, Web sentiment analysis for scoring positive or negative words using Tweeter data. *Int. J. Comput. Appl.* **96**(6), 33–37 (2014)
28. A Hogenboom et al., Lexicon-based sentiment analysis by mapping conveyed sentiment to intended sentiment. *Int. J. Web Eng. Technol.* **9**, 125–147 (2014)
29. RF Martins, A Pereira, F Benevenuto, "An Approach to Sentiment Analysis of Web Applications in Portuguese." *Proceedings of the 21st Brazilian Symposium on Multimedia and the Web ACM*, 2015, pp. 105–112
30. C Clavel, Z Callejas, Sentiment analysis: from opinion mining to human-agent interaction. *IEEE Trans. Affect. Comput.* **7**(1), 74–93 (2016)
31. Z Xu et al., Knowle: a semantic link network based system for organizing large scale online news events. *Future Generation Comput. Syst.* **43–44**, 40–50 (2015)
32. Z Xu et al., Crowdsourcing based social media data analysis of urban emergency events. *multimedia tools Appl.* DOI:10.1007/s11042-015-2731-1. (2015)
33. Z Xu et al., Crowdsourcing based description of urban emergency events using social Media big data. *IEEE Trans. Cloud Comput.* DOI:10.1109/TCC.2016.2517638.(2016)
34. Z Xu et al., Participatory sensing based semantic and spatial analysis of urban emergency events using mobile social media. *EURASIP J. Wireless Commun. Netw.* **44**, 2016 (2016)
35. Z Xu et al., Building knowledge base of urban emergency events based on crowdsourcing of social media. *Concurrency Comput. Pract. Exp.* DOI:10.1002/cpe.3780. (2016)

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Immediate publication on acceptance
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► springeropen.com