Journal of Computers Vol. 28, No. 3, 2017, pp. 159-173
doi:10.3966/199115592017062803013

Micro-blog Opinion Leader Selection Method Using Emotional Contribution Model

Ya-Xing Li[1], Zhao-Kai Wang[1], Li-Jun Liu[1], Xu-Peng Feng[2], Li Liu[1], Qing-Song Huang[1,3*]

[1] Faculty of Information Engineering and Automation, Kunming University of Science and Technology
Kunming 650500, China
{lyx02180607, sanxianwei}@163.com
{93612798, 48240664}@qq.com

[2] Educational Technology and Network Center, Kunming University of Science and Technology
Kunming 650500, China
417062815@qq.com

[3] Yunnan Key Laboratory of Computer Technology Applications, Kunming 650500, China
kmustailab@hotmail.com

Received 30 May 2016; Revised 09 December 2016; Accepted 09 December 2016

Abstract. To solve the false correlation caused by negative influence in selecting opinion leader, a micro-blog opinion leader selection method using emotional contribution model is proposed. When the retweeters and reviewers of the tracked object is made, there is not all positive content but kinds of contents of negative, thus they are divided into effective influence or ineffective, which means the all contents made by retweeters and reviewers are taken as an emotional contribution model. In the process of object tracking by using emotional contribution model, messages dissemination occurs mostly in the early, so the early spread the greater influence. As the result, when the tracked object has negative forwarding or reply, the proposed method can infer the polarity of emotion based on variant LSTM. Then, we get the effective influence ranking of opinion leader by each object's coverage rate. The experimental results show coverage rate of proposed model has improved 4.9% than the Page Rank algorithm.

Keywords: coverage area, message dissemination, opinion leader, sentiment classification

1 Introduction

Opinion leader selection has become a hot issue with the rapid development of computer, particularly in applications such as the guidance and development of public opinion, promotion of products and corporate propaganda. However, opinion leader selection is extremely complex and time consuming. Some contents said by retweeters and reviewers are negative to tracked object. To overcome these difficulties, an emotional contribution model has been proposed. It effectively removes the user whose content has a negative impact or exceeds the specified time. Micro-blog's influence has reached the extent of other medias can't replace [1-2]. Forwarding behavior decides whether or not the object has chance to be viewed by others in micro-blog [3].

In recent years, some scholars combine users' attribute value, network structure and diffusion capacity to select opinion leaders. Huang [4] proposed an algorithm Discovering Network Community, which constructs objective function with modularity Q, Cut and silhouette. Tian [5] put forward a new user modeling method which mixed traditional link analysis and user attributes similarity together, effectively identify the topic high influence user. Fan [6] presented a new Influence Diffusion Probability Model

* Corresponding Author

(IDPM), and then builds a network opinion leader identification model. Wu [7] presented an algorithm of topic-related opinion leaders mining. First, a topic-related graph model of micro-blog was built according to users' attributes and mutual information among them. Then the idea of random walk was adopted in the algorithm to mine micro-blog opinion leaders by finding central nodes of the graph model. Ding [8] presented a method that combined random walk with filter probability method from internal across network to mine individual of topic influence aimed at multiple network.

There are three main methods of mining Opinion leaders in social network: First, calculating the user's own attribute value [7] or by clustering to find the opinion leader; Second, according to the relationship between the users in the social network, authors got the importance of users in the network structure based on the PageRank algorithm to mine the opinion leaders [4-5]; Third, authors calculated the user's importance from the perspective of information diffusion [6]. But, literatures [4-8] neglected the emotional tendency of all users. In this paper, we think that the emotion between the users is of great significance to the opinion leader. And for SINA micro-blog, its contents update fast, one of important thing need to consider is that SINA micro-blog sends contents to users to browse according to the time sequence, so it is necessary for users to interact with opinion leader in time, which can effectively enhance the influence of opinion leaders, or the influence that the user passes to the opinion leaders is a little weaker.

Text sentiment classification is mainly based on the method of emotional knowledge and feature classification [9]. Turney and Littman [10] used Mutual Information-Information Retrieval Point-wise (PMI-IR) method to calculate the correlation between words and seed words. Zhu [11] presented two methods including semantic similarity and semantic field to calculate the correlation between words and appraise benchmark words drawn support from HowNet ontology library. However, traditional sentiment classification needs to extract the sentiment dictionary in advance, which means the accuracy depends on manual operation, and the result is just passable. Deep learning allows establishing the model under the premise of "zero dictionary".

In conclusion, the contribution of this paper comprises of the following components: (1) We used recurrent neural network to build emotion classification model based on LSTM, and got the users' emotion tendency. (2) Microblogging as the short text, it has a large amount of information and its contents update fast, so opinion leaders' influence including that people who influenced by opinion leader will interact with him or her timely not after a few days. So, in the experiment part, we analyzed the rising cycle of topics' heat to get the effective interaction time. (3) This paper proposed an emotional contribution model, and we removed the negative influence, and message dissemination is from a node to other nodes in our model in the effective time range. On the one hand, in spite of simple contents, the method analyzes the polarity of emotion accurately. On the other hand, the effective time range can be estimated from visible statistics. (4) Lastly, we used Breath-First Search (BFS) algorithm whose node depth is two to get the user' node coverage.

The rest of the paper is organized as follows. In section 2, framework of opinion leader selection method is briefly introduced. Following that, emotional contribution model is proposed and built in section 3. In section 4, the proposed method is described with micro-blog platform. In addition, experimental results are presented in section 5. Finally, the conclusions are drawn in section 6.

2 Essentials of Opinion Leader Selection Method

The opinion leader selection method mainly describes the development of the transitive relation between users related with object. Object is a message released by a user. The method can be represented by a contribution diagram, which is constructed by users' attribute values and weight. Attribute value is depend on Loyalty and the number of fans. Loyalty is a loyal degree of retweeters and reviewers to publishers. The weight relies on historical forwarding probability and users' activity. Activity is a micro-blog exposure rate of publisher to retweeters. The contribution diagram is simplified by the forwarding time and the emotional polarity of retweeter. The effective time range is that forwarding time can't exceed time threshold CF. The positive influence of a retweeter is that incidental content's emotional polarity is positive.

Message dissemination in micro-blog includes publishers and retweeters. We define transitive relation in contribution diagram is a two tuple C = (pub, pro) consisting of the publisher pub and the set of retweeters pro = $\{U_1, U_2, \ldots, U_n\}$. The process of publishers releasing message to retweeter is denoted as

pub→pro. A. retweeter belonging to transitive relation C is a publisher belonging to transitive relation D, we name it as a development denoted as <C|D>.

Lan is the retweeters' contents edited in transitive relation, whose emotional polarity is denoted as *EMO* (*Lan*) with value range [-1, 1]. If *EMO* (*Lan*) is positive, it indicates that the spread of the retweeter enhances the influence of the publisher. If not, remove the relation.

c_time is the time when a retweeter forwards a message from publisher, who releases the message at *f_time*, the more similar *c_time* and *f_time*, the greater influence of the publisher. According to the *c_time* and polarity of *EMO* (*Lan*), we judge the validity of developments recursively to simplify contribution diagram, thus getting the effective opinion leaders ranking on the basis of users' coverage rate.

3 Opinion Leader Selection Method Based on Emotional Contribution Model

User's influence is the capacity that a user releases a message and inspires another users forwarding. In the conventional method, selecting the opinion leader often neglects emotional analysis and the factor of time. In our proposed method, the opinion leaders are obtained from the contribution diagram. Message dissemination is from a node to other nodes in the diagram. By this means, the method can obtain the forwarding time and analyze the emotional polarity between two users. In this section, opinion leader selection method based on emotional contribution model is explained.

$A(i) = \{m \mid (i, m) \in G\}$ is a user set focused by user i, $Ne(m) = \{i \mid (i, m) \in G\}$ is a set of fans belonging to user m. G is a set of all users. $W = \{W_1, W_2, \ldots, W_n\}$ is all users' micro-blog set, which $W_i = \{W_{i1}, W_{i2}, \ldots, W_{iNum_i}\}$, $1 \le i \le n$, $1 \le j \le Num_i$.

3.1 Node Value Calculation

Node value is a user's attribute value, made of *Loyalty* and the number of fans. Effective history forwarding number and reply number reflects a loyal degree donated as *Loyalty*. *Loyalty* is also depending on emotional analysis and forwarding time. The more and effective replies and forwarding and the faster speed of replies and forwarding, the more probably the message dissemination can be appeared between users.

3.1.1 Emotional Analysis

Many data can be got from corpus such as historical forwarding number, the number of replies and forwarding time, however the number of positive replies depends on the emotional polarity. If the polarity of replies is positive, it's effective, otherwise it is invalid. Traditional sentiment classification needs to extract the sentiment dictionary in advance, that means the accuracy depends on manual operation, and the result is just passable. However, deep learning allows establishing the model under the premise almost "zero dictionary". We present a text sentiment classification model based on variant LSTM, whose structure is shown in Fig. 1.

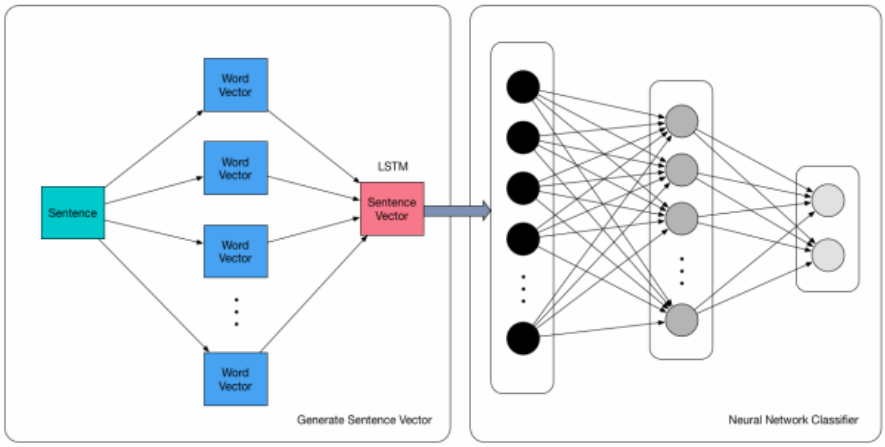

Fig. 1. Structure of sentiment classification

The structure of sentiment classification makes collection of replies as input, the emotional polarity as output. First of all, we convert sentences to term vectors which are the input of variant LSTM. Then, the output of the variant LSTM is used as the input of the neural network classifier. Finally, we get the sentimental polarity through the hidden layer. LSTM's internal structure has three layers, as shown in Fig. 2.

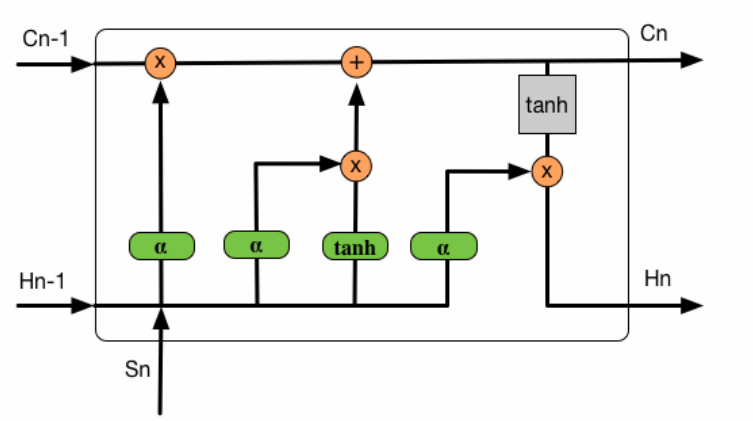

Fig. 2. Internal structure of LSTM

LSTM is consisted of input gate, output gate, forgotten gate and memory cells. The first three gates are controllers to control writing, reading and loss operations of memory cells, protecting and controlling the cell states. The gate is a method that information selects the way to pass. A formalized representation of LSTM is as follows:

$$f_t = \sigma(W_f \cdot s_n + U_f \cdot h_{n-1} + V_f \cdot c_{n-1} + b_f) \quad \textbf{(1)}$$

$$i_t = \sigma(W_i \cdot s_n + U_i \cdot h_{n-1} + V_i \cdot c_{n-1} + b_i) \quad \textbf{(2)}$$

$$c_t = f_t \times c_{n-1} + i_t \times \tan h(W_c \cdot s_n + U_c \cdot h_{n-1} + b_c) \quad \textbf{(3)}$$

$$o_t = \sigma(W_o \cdot s_n + U_o \cdot h_{n-1} + V_o \cdot c_{n-1} + b_o) \quad \textbf{(4)}$$

$$h_n = o_t \times \tan h(C_t) \quad \textbf{(5)}$$

Where σ is the activation function, W_*, U_*, V_*, and b_*, denote the coefficient matrix and bias vector respectively, i_t, f_t and o_t denote the calculation methods of input gate, output gate and forget gate respectively at time t. c_t denotes a calculation method of forgotten gate at time t. h_n denotes the result of LSTM.

Variant LSTM's internal structure is shown as in Fig. 3. A formalized representation of variant LSTM is as follows:

$$f_t = \sigma(W_f \times [h_{n-1}, +s_n] + b_f) \quad \textbf{(6)}$$

$$\hat{c}_t = \tan h(W_c \times [h_{n-1}, s_n] + b_c) \quad \textbf{(7)}$$

$$c_t = -f_t \times C_{t-1} + (1 - f_t \times \hat{c}_t \quad \textbf{(8)}$$

$$o_t = \sigma\ (W_o \times [h_{n-1}, s_n] + b_o) \quad \textbf{(9)}$$

$$h_n = o_t \times \tan h(C_t) \quad \textbf{(10)}$$

$$\sigma = \frac{1}{1+e^{-x}} \tag{11}$$

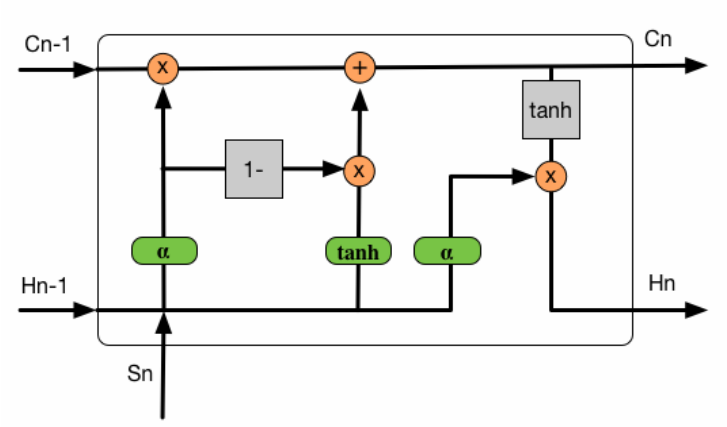

Fig. 3. Internal structure of variant LSTM

In variant LSTM, we don't need to ensure when to forget some information and what's time to add some new information. OC represents a set of original replies. OCP represents the polarity of replies. The specific algorithm is described as follows:

Algorithm1: Feature Extraction

```
Input: OC
Output: h*
Variable: h_{n-1}, f_t, i_t, o_t, ĉ_t, h_n
Feature Extraction (OC) {
//Convert the OC into term vector(200 dimensions)
// Transform word vector into sentence vector as the input of
variant LSTM
Deciding what information is discarded and adding information at
the same time{
   f_t is discarded
   tan h creates a new candidate vector ĉ_t
   C_t = f_t × C_{t-1} + (1 − f_t) × ĉ_t
}
Output value{
   o_t is the output of cell state
   h_t = o_t × tan h (C_t)
}
RETURN h_t
}
```

In neural network classifier, it has three layers: input layer, hidden layer and output layer. h_i is the input and output of i unit in input layer. net_j is the input of j unit in hidden layer. Y_j is the output of ^{j} unit in hidden layer. net_k is the input of k unit in output layer. k Z_k is the output of k unit in output layer.di, dh and do denote the number of units in input layer, hidden layer and output layer, they are 256, 128 and 2.

Algorithm2: Neural Network Classifier

```
Input: h*
Output: OCP
```

Variable: net_j, ω_*, net_k, Y_*, Z_*

Neural Network Classifier (h_*) {

//Input layer

Activation function is $f(x)=\begin{cases}\gamma & x \geq \gamma \\ x & |x|<\gamma \\ -\gamma & x<-\gamma\end{cases}$ **(12)**

//is the maximum value in input layer

The output of i unit in input layer is h_i

//Hidden layer

Activation function is $f(x)=\frac{1}{1+e^{-x}}$ **(13)**

The input of j unit in hidden layer is net_j

$$net_j=\sum_{i=1}^{di}\omega_{ji}*h_i+\omega_{j0}=\omega_j^t h \quad (14)$$

$$h=(h_0,h_1,...,h_{di})^t \quad (15)$$

$$h_0=1 \quad (16)$$

$$\omega_j=(\omega_{j0},\omega_{j1},...,\omega_{jdi})^t \quad (17)$$

The output of j unit in hidden layer is Y_j

$$Y_j=f(net_j) \quad (18)$$

//Output layer.

Activation function is $f(x)=\frac{1}{1+e^{-x}}$

The input of k unit in output layer is net_k

$$net_k=\sum_{j=1}^{dh}\omega_{kj}*Y_j+\omega_{k0}=\omega_k^t Y \quad (19)$$

$$Y=(Y_0,Y_1,...,Y_{dh})^t \quad (20)$$

$$Y_0=1 \quad (21)$$

$$\omega=(\omega_{k0},\omega_{k1},...,\omega_{jdh})^t \quad (22)$$

The output of k unit in output layer is Z_k

$$Z_k=f(net_k) \quad (23)$$

}

3.1.2 Impact of Forwarding Time

The speed of forwarding time is important to the message dissemination, the earlier dissemination, and the better effect. Lee et al. [13] got the number of twitter users related the current topic and observed the number changing with time. We obtain number of forwarding changes with time for four topics.

It can be seen that the growth rate of the forwarding number is very fast at the beginning from Table 1, with the passage of time, the growth rate tends to be slow. This behavior shows that message dissemination occurs mostly in the early, the earlier dissemination, and the better effect. That means the influence created by a retweeter who forwarded early better than those retweeters who forwarded more late. The effective time range is that forwarding time can't exceed time threshold CF, and δ_{CF} is time coefficient in the effective time range. If forwarding probability is less than $Min(\delta_{CF})$, the corresponding edge in contribution diagram needs to be pruned.

Table 1. Change of forwarding volume

Topic	First Day	Second Day	Third Day	Fourth day
Typhoon of "Canhong"	3156	2734	785	53
Tianjin explosion	173192	186539	47902	9208
Japan surrender	30575	25165	6501	1731
Forever Young	159572	177517	53860	7981

3.2 Weight Calculation

According to section 3.1, Defining fan_i is the number of Ne(i), Att_i is the number of A(i), m is the historical replies number of messages released by user i, n is the history forwarding number, p is the valid replies number and q is the number of effective forwarding. Loyalty is shown as formula (24).

$$\text{Loyalty} = \frac{p+q}{m+n} \tag{24}$$

the influence of node itself is U_i_Inf .

$$U_i\_\text{Inf} = \text{Loyalty} * fan_i \tag{25}$$

the weight between nodes is the forwarding probability EC(i, m).

We define that EC(i, m) is Re(i, m) multiplies by Len(m, i). Re(i, m) is the history forwarding probability from i to m. Len(m, i) is the exposure rate of messages from m toi. The formula is shown as (26)

$$\text{Ec}(i, m) = \text{Re}(i, m) * \text{Len}(m, i) \tag{26}$$

Sub formulas are as follows:

$$Len(m,i) = \frac{NuW_m}{\sum_{\tau \in A(i)} NuW_\tau} \tag{27}$$

NuW_m is the number of messages released by user m, $\sum_{\tau \in A(i)} NuW_\tau$ is the number of all messages released by A(i).

$$Re(i,m) = \frac{NuR[i,m] + \frac{1}{Att_i}}{\sum_{\tau \in A(i)} NuR[i,m]} \tag{28}$$

NuR[i, m] is the forwarding number from user m to i. The specific algorithm described as algorithm 3.

Algorithm3: Forwarding Probability

Input: $A(i) = \{m \mid (i,m) \in G\}$, Att_i, $W_i = \{W_{i1}, W_{i2}, ..., W_{iNum_i}\}$

Output: EC

Variable: yh, $\text{Re}(i,m)$, $\text{Len}(m,i)$, $\text{Pro}(i,m)$

Forward Calculate(A(i), W_i, Att_i){

$\text{NuR}[i,m]$ is a statistics that the number of forwarding from i to m

NuW_m is a statistics that the number of messages released by m

FOR EACH yh IN A(i) {

$\sum_{\tau \in A(i)} \text{NuR}[i,\tau]$ is a statistics that the number of user i forwards user A(i)'s messages

$\sum_{\tau \in A(i)} NuW_\tau$ is a statistics that the number of messages released by A(i) .

}

$$Len(m,i) = \frac{NuW_m}{\sum_{\tau \in A(i)} NuW_\tau}$$

$$Re(i,m) = \frac{NuR[i,m] + \frac{1}{ATT_\tau}}{\sum_{\tau \in A(i)} NuR[i,\tau]}$$

$$EC(i,m) = Re(i,m) * Len(m,i)$$

RETURN EC(i, m)

}

A = {a, {b, c}}

3.3 Construction and simplification of Contribution diagram

The publishers and retweeters in transitive relations are nodes, and forwarding probability is weight of edges in contribution diagram. Giving an example, transmission relations as follows: $A = \{a, \{b, c\}\}$, $B = \{b, \{d, e\}\}$, $C = \{c, \{f, g, h\}\}$, $D = \{d, i\}$, $E = \{f, j\}$, in which the development set has <A|B>, <A|C>, <B|D>, <C|E>. The initial structure of the contribution diagram is shown in Fig. 4.

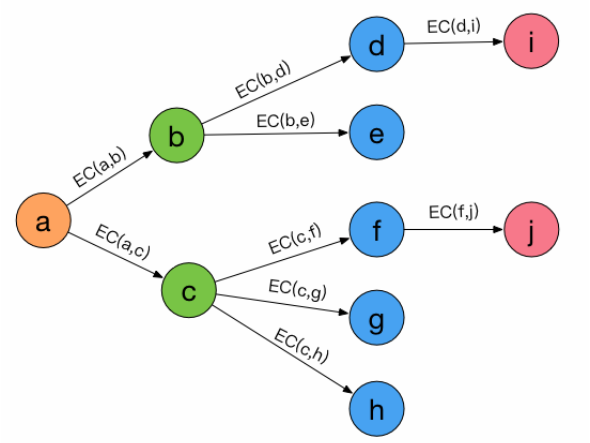

Fig. 4. Initial contribution

Node value is U_*Inf. Forwarding message set is denoted as $rep = 1, 2, ..., M$. $Re_{rep,q}$ shows the rep is forwarded by q times, which q means the number of times of message to be forwarded. The forwarding time is c_{time}. Lan is the retweeters' contents edited during forwarding. $Re_{rep,0}$ is the initial message.

In $Re_{rep,q-1} \rightarrow Re_{rep,q}$, if EMO (Lan) edited by U_{ie} is positive, it indicates that the spread of the retweeter enhances the influence of the publisher, the influence of fans is positive. If not, remove the relation. The specific algorithm described as algorithm 4. trs is the transitive set, trss is the transitive set after simplification.

Algorithm4: Contribution Diagram Simplification

```
Input: trs
Output: trs, EC
Variable: CF = c_time – f_time Pra = EMO * δ_cf
DevelopmentGraph Simplify (trs) {
FOR EACH node IN trs {
  IF node has no publisher{
    Forward Calculate( A(i), W_i , Att_i )
    //Predict retweeter of the node and form the transfer relation
```

```
        Feature Extraction (OC)
        Neural Network Classifier (h*)
        //Obtain forwarding time.
         IF  Pra < −Min(δcf)  {
          Remove the edge of the reteeeter and publisher
         }
         IF node is conclude{
           REMOVE
         }
         DevelopmentGraph Simplify(trs)
    }
}
```

Via calculating the emotional polarity, forwarding probability, forwarding time and $U_i\text{Inf}$, We can get the simplified Fig. 5 by Fig.4 through evolving the transitive relation B and E.

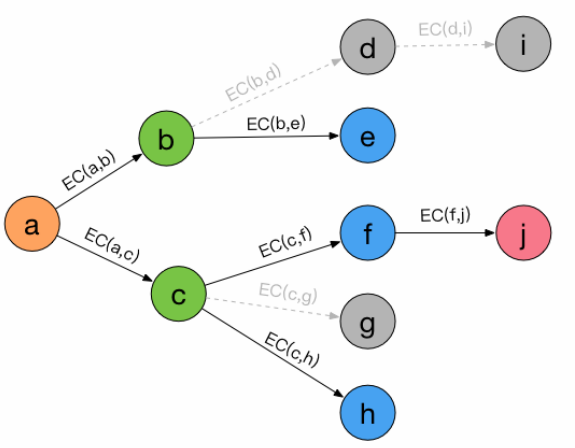

Fig. 5. Simplified contribution

4 Micro-blog Opinion Leader Selection

In this section, the proposed method is described with micro-blog platform. Message dissemination mainly depends on retweeters and reviewers in micro-blog platform. This paper expresses the micro-blog opinion leader selection method like a contribution diagram. The contribution diagram contains nodes with $U_*\text{Inf}$ and edges with forwarding probability, is a core of emotional contribution model, which can express the recursive process of selecting opinion leader intuitively. Fig. 6 is a process diagram.

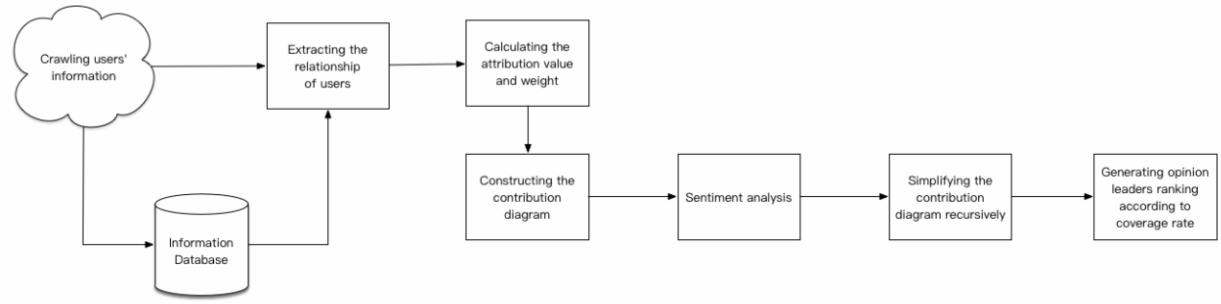

Fig. 6. Process of micro-blog opinion leader selection

The micro-blog opinion leader selection method based on emotional contribution model firstly obtains experimental data by crawling from micro-blog platform. Then, calculating each user's attribute value and forwarding probability between two users, to construct the contribution diagram. Next according to sentiment analysis and forwarding time to simplify the diagram. Finally we get the ranking of opinion

leader through coverage rate of every effective node.

5 Experimental Results

In this section, the experimental results are presented by three groups of experiments to illustrate the performance and the effectiveness of the proposed method. The first set of experiments demonstrates the time threshold selection and time coefficient in the effective time range. The second set of experiments compares the accuracy of sentiment analysis of the three algorithms in different number of Feature Size, and the third set of experiments compares the node coverage rate of opinion leader of three methods in different number of opinion leader. In these experiments, the experimental data includes the user and messages released in micro-blog platform from 2015 March to October 2015.

According to the characteristic of opinion leaders can affect a large number of other users [16]. We define the coverage rate is as follows.

$$P(u) = \frac{\sum_{u=1}^{N} Numer(u)}{N} \tag{29}$$

$P(u)$ is the coverage rate of top u, N is the number of all users, $Numer(u)$ is the number of affected user by top u.

Accuracy rate and the recall rate as the evaluation criterion in sentiment classification.

5.1 Threshold Setting

The speed of forwarding time is important to the dissemination of users' influence. The experiment shows the forwarding number change with time, as shown in Fig. 7

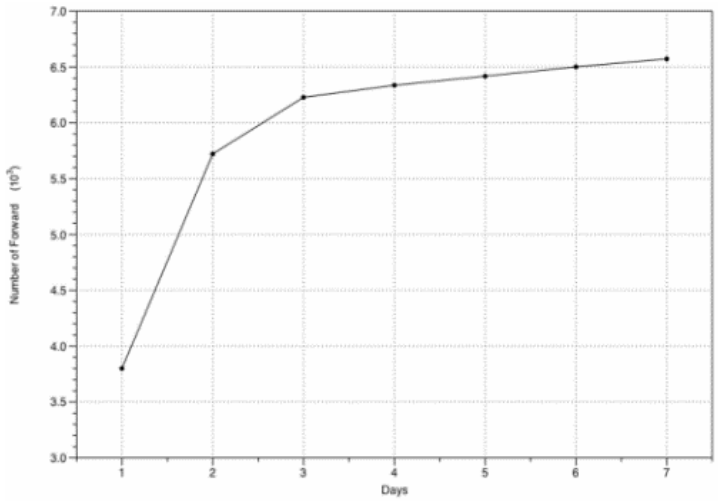

Fig. 7. Forwarding number change with time

With the change of time, the growth rate of the forwarding number of is getting lower and lower and gradually tends to 0. The fastest growing time is in the first three days, so $CF = 4$ is a limit. We define the effective forwarding time is the first three days.

Fig. 8 shows the number of forwarding for four topics change with the forwarding time that no more than CF.

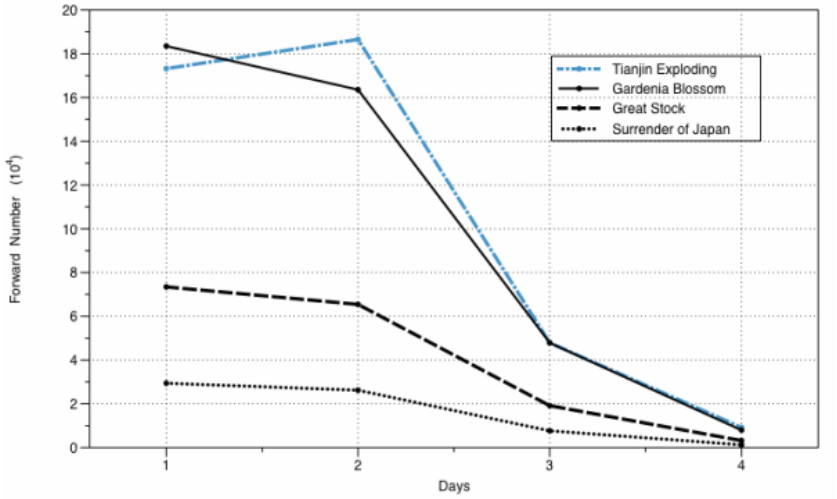

Fig. 8. Change of forwarding number

It can be seen from Fig. 8 that blue marked is different from the other three, whose forwarding number is on the rise in first two days, it's because that the incident involved in life safety, so it got a wide range of attention. However, the rate of change for the other three topics is basically similar. The time coefficient is defined as δ_i, i represents the number of days that between c_{time} and f_{time}. According to the trend of the number of forwarding change over with time in the Fig. 8, set the $\delta_1 = 0.46$, $\delta_2 = 0.41$, $\delta_3 = 0.12$ after comparing the proportions synthetically.

5.2 Emotional Classification

The accuracy rate P and recall rate R are the evaluation criteria in this section. P and R are calculated as follows:

$$P = \left(\frac{a}{a+b} + \frac{d}{c+d} \right) / 2 \tag{30}$$

$$R = \left(\frac{a}{a+c} + \frac{d}{b+d} \right) / 2 \tag{31}$$

$$F = \frac{2 * P * R}{P + R} \tag{32}$$

a is the number of texts which is known polarity is positive determined polarity is positive, b is the number of texts which is known polarity is negative determined polarity is positive, c is the number of texts which is known polarity is positive determined polarity is negative, d is the number of texts which is known polarity is negative determined polarity is negative.

Experiments were performed with different feature sizes including 500 and 2000 and different methods including Common-SVM and Mixed-SVM. The results of three methods are compared in Fig. 9 and Fig. 10.

The accuracy rate and recall rate change with feature size, the larger feature size, the higher P and R. It can be seen that results are effective in using variant LTSM network to analyze emotional polarity. The accuracy rate of the proposed model increases 4.3% than Mixed-SVM, which compared with the Common-SVM, the accuracy rate has improved 7.5%.

SVN is a relatively mature classification algorithm, Morris et al. [19] proposed a classification method based on different features is proposed (MCP). In order to fully verify the validity of the emotions classification based on LSTM, we selected the proposed classification method to compare respectively with the baseline algorithms SVM and MCP proposed in Morris et al. [19] in the accuracy and recall rate. The experimental results are shown in Table 2.

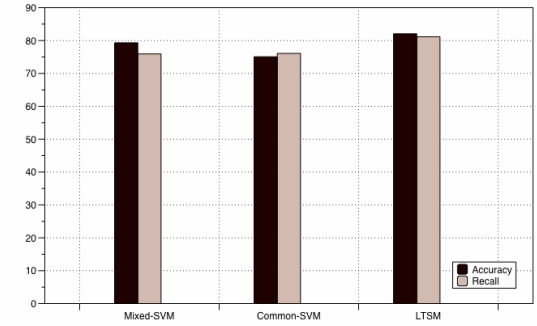

Fig. 9. Feature Size is 500

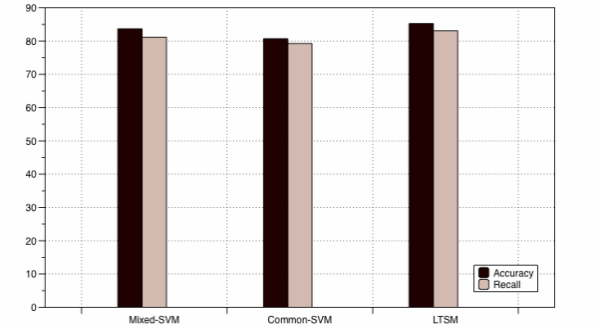

Fig. 10. Feature Size is 2000

Table 2. Comparison results of different methods (Feature Size is 500)

Method	P	R	F
SVM	0.7197	0.7443	0.7318
MCP	0.791	0.79	0.7905
LSTM	0.8125	0.7970	0.8047

From the Table 2, we can see that the accuracy rate, recall rate and comprehensive index value of traditional SVM classification methods is about a little more than 0.70, and the method based on the different features and LSTM model can effectively determine the emotional tendency of micro-blog contents. Huang et al. [4] uses English corpus to make the experiment and we used the Chinese language as the corpus in this experiment, therefore, the experimental results of MCP has a certain extent deviate from the original paper. At the same time, it is proved that the proposed method is effective.

5.3 Opinion Leaders Ranking

5.3.1 Coverage Area Comparison

It is believed that a stable UserRank value can be obtained after multiple access procedure probability transfer between the users which was similar to PageRank algorithm [5]. A WIR model is proposed [18], whose evaluation indicator is the covering number in propagation process of each opinion leader. So in this paper, we selected the proposed model to compare respectively with the baseline algorithms PageRank and WIR proposed in the covering number [18]. The experimental results are shown in the following Fig. 11.

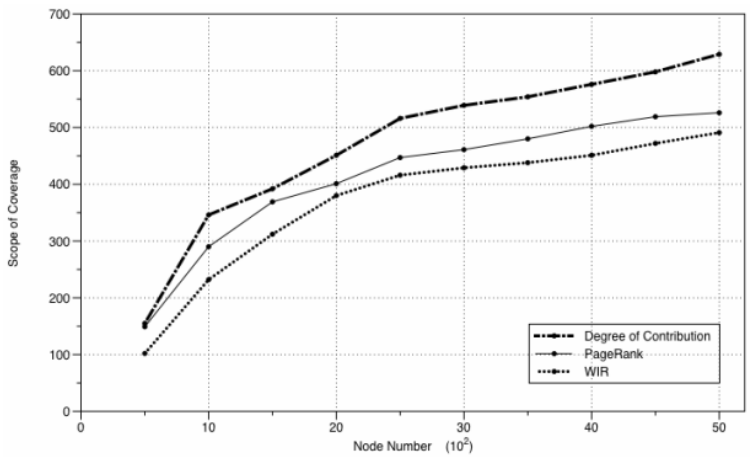

Fig. 11. Comparison of coverage area of different methods

Fig. 11 shows the PageRank algorithm is similar to the proposed model at the beginning, with the increase of the node number, the gap is getting bigger. This is because the PageRank algorithm takes node degree as the core measure. The proposed model cares about coverage area. The coverage area of

the WIR algorithm tends to be slow with the number of opinion leaders increased, it's because the first N Node gradually polymerization in WIR algorithm.

5.3.2 Accuracy Comparison

5.3.1 Part has been got the opinion leaders coverage area ranking. However, how many of these opinion leaders are true opinion leaders need to be tested. Fan et al. [6] used P@N as the evaluation indicator, which used artificial judgement to find the real opinion leaders in the rankings. In this paper, we let three teachers in our experimental team to test these opinion leaders that whether these opinion leaders are true opinion leaders. In this part, we compared proposed method with literature [6]. The experimental results are shown in the following Table 3 to Table 6.

Table 3. Accuracy comparison results (N = 5)

Method	Average Correct Rate /%	Change Rate Compared with OL_IDM /%	Change Rate Compared with OL_IDPM /%
OL_IDM	19.6	-	-
OL_IDPM	91.0	+364.3	-
Proposed Method	96.0	+389.8	+5.5

Table 4. Accuracy comparison results (N = 10)

Method	Average Correct Rate /%	Change Rate Compared with OL_IDM /%	Change Rate Compared with OL_IDPM /%
OL_IDM	34.1	-	-
OL_IDPM	88.0	+158.1	-
Proposed Method	91.5	+168.3	+3.98

Table 5. Accuracy comparison results (N = 20)

Method	Average Correct Rate /%	Change Rate Compared with OL_IDM /%	Change Rate Compared with OL_IDPM /%
OL_IDM	37.3	-	-
OL_IDPM	76.8	+105.9	-
Proposed Method	86.4	+131.6	+12.5

Table 6. Accuracy comparison results (N = 50)

Method	Average Correct Rate /%	Change Rate Compared with OL_IDM /%	Change Rate Compared with OL_IDPM /%
OL_IDM	43.3	-	-
OL_IDPM	71.2	+64.4	-
Proposed Method	81.5	+88.2	+14.5

From Table 3 to Table 6, we can see that the accuracy rate value of OL_IDM methods is a little low, method OL_IDPM and based on emotional contribution model can effectively mine the opinion leaders. Fan et al. [6] uses post and its comments to make the experiment and we used the SINA micro-blog that is famous for short text as the corpus in this experiment, therefore, the experimental results of literature [6] has a certain extent deviate from the original paper. At the same time, it is proved that the proposed method is effective.

6 Conclusions

To solve the false correlation caused by negative influence in selecting opinion leader, the micro-blog opinion leader selecting method and emotional contribution model are studied, and opinion leader selection is applied to emotional contribution model. To realize reliable analysis of emotional, especially when the forwarding and replies are in any time including invalid time, we use emotional contribution model to find the negative influence, and simplify the contribution diagram combined with invalid time.

Experimental results have demonstrated that the proposed micro-blog opinion leader selection method is effective.

Due to the contents are colloquial in real social network and the relationships between users are complex, so these problems have brought many limitations and challenges to our research, the details are as follow: (1) Limited to SINA API, we are unable to get access to all micro-blogs of one person and the comments and forwarding of the micro-blogs. Therefore, due to lack of large-scale corpus, the accuracy needed to be further strengthened. (2) There is some "Navy" or "Zombie powder" in user set selected in the paper, which are very active in the community, so the accuracy of mining opinion leaders has been influenced. (3) Due to the texts of micro-blog are not only informative, but also renewal contents quickly. The computational efficiency is severely restricted.

To the above problems, what we should do first is to collect and obtain more data sets as much as possible. Second, one critical issue in the optimization algorithm is how to introduce surveillance mechanism for effective monitoring on the "Navy" or "Zombie powder". Third, we need to explore distributed opinion leader identification algorithm based on cloud computing platform in future research work, which can improve its computational efficiency and enhance its value of enterprise application.

Acknowledgment

This work was supposed by the National Nature Science Foundation of China (81360230, 81560296, 61462056, and 61462051).

References

[1] H. Kwak, C. Lee, H. Park, S. Moon, What is twitter, a social network or a news media? in: Proc. the 19th WWW. Raleigh: ACM Press, 2010.

[2] J. Jansen, M. Zhang, K. Sobel, A. Chowdury, Twitter power: tweets as electronic word of mouth, Journal of the American Society for Information Science and Technology 60(11)(2009) 2169-2188.

[3] L. Zou, Social network analysis based on micro-blog communication mechanism, Seeker 11(2013) 241-243

[4] F. Huang, S. Zhang, X. Zhu, Discovering network community based on multi-objective optimization, Journal of Software 24(9)(2013) 2062-2077.

[5] Tian X., Song Y., Zhu t, et al. Effective community Leader election method based on user similarity measure, Journal of Yanshan University 38(6) (2014) 516-521.

[6] X.-H. Fan, J. Zhao, B.-X. Fang, Y.-X. Li, Influence diffusion probability model and Utilizing it to identify network opinion leader, Journal of Software 36(2)(2013) 360-367.

[7] X.-H. Wu, H. Zhang, C.-M. Yang, B. Li, X.-J. Zhao, An algorithm of topic-related microblogging opinion leader mining, Journal of Chinese Computer Systems 35(10)(2014) 2296-2301.

[8] Z.-Y. Ding, Y. Jia, B. Zhou, Y. Han, Mining topic influencers based on the multi-relational network in micro-blogging sites, China Communications 1(2013) 93-104.

[9] Y.-Y. Zhao, B. Qin, T. Liu, Sentiment analysis, Journal of Software 21(8)(2010) 1834-1848.

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

[11] Y.-L. Zhu, J. Min, Y.-Q. Zhou, Z.-J. Huang, L.-D. Wu, Semantic orientation computing based on HowNet, Journal of Chinese Information Processing 20(1)(2006) 14-20.

[12] J. Liang, Y. Chai, H. Yuan, M. Gao, H. Zan, Polarity shifting and LSTM based recursive networks for sentiment analysis, Journal of Chinese Information Processing 29(5)(2015)152-158.

[13] C. Lee C, H. Kwak, H. Park, S. Moon, Finding influential based on the temporal order of information adoption in Twitter, in: Proc. the 19th International Conference Companion on world wide Web (WWW'10), 2010.

[14] K. Yue, C.-L. Wang, Y.-L. Zhu, H. Wu, W.-Y. Liu, Click-through rate prediction of online advertisements based on probabilistic graphical model, Journal of East China Normal University (Natural Science) 3(2015) 15-25.

[15] C. Xiong, T. Wang, W. Ding, Y. Shen, T.-Y. Liu, Relation click prediction for sponsored Search, in: Proc. WSDM'12, 2012.

[16] Y. Wu, L. Ma, M. Lin, H.-T. Liu, Discovery algorithm of opinion leaders based on user influence, Journal of Chinese Computer Systems 36(3)(2015) 561-565.

[17] Z. Zhai, H. Xu, P. Jia, Identifying opinion leaders in BBS, in: Proc. 2008 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology, 2008.

[18] K. Wu, X. Ji, J. Guo, C. Liu, Influence maximization algorithm for micro-blog network, Journal of Computer Applications 33(8)(2013) 2091-2094.

[19] M.R. Morris, S. Counts, A. Roseway, A. Hoff, J. Schwarz, Tweeting is believing? understanding microblog credibility perceptions, in: Proc. the 15th ACM Conf on Computer Supported Cooperative Work (CSCW12), 2012.

[20] T. Ge, C. Xue Chuanye, The study of SINA micro-blog opinion leaders effect on network public opinion transmission, International English Education Research 3(2015) 68-72.