Eurasia Journal of Science and Technology	ISSN: 2663-1024	Volume 2, Issue 2,pp 64-70	April 2020

SENTIMENT TREND PREDICTION OF MICROBLOG USERS BASED ON MULTI-LEVEL ATTENTION MECHANISM

Cheng Zhao[1,2,*]

[1]*Command and Control Engineering College, Army Engineering University of PLA, Nanjing 212000, Jiangsu, China*

[2]*School of Computer and Information, Anhui Normal University, Wuhu 241000,Anhui, China*

**E-mail address: ntmaple@mail.ahnu.edu.cn*

Abstract

In order to improve the accuracy of social network users' emotional behavior prediction, this paper proposes a multi-level attention mechanism based microblog user group emotional trend prediction. Based on Sina Weibo data, the multi-level attention mechanism is used to divide the text content emotion of user groups and collect the emotional characteristics of user groups. This paper analyzes the group emotional behavior of microblog users and establishes a hierarchical emotion processing model. The model of emotion transfer of microblog users is constructed, and the problem of unstable convergence value is solved by direct fuzzy algorithm. Using the node matching method to measure the emotional similarity, the sentiment trend of microblog users is predicted. Taking the microblog data from November to December in 2019 as an example, the experimental results show that the emotional trend predicted by this method is close to the actual trend, which shows the effectiveness of the method in this paper.

Keywords: multi level attention; microblog users; group emotional trend; feature collection

1. INTRODUCTION

Social network emotional behavior analysis, especially group social network emotional behavior analysis, has important application value for public opinion monitoring [1]. In recent years, although the empirical research and theoretical research on group emotional behavior in social network has made certain progress, the discussion on group emotional behavior is still in the initial stage, and there are still many questions worthy of in-depth study, mining and waiting for answers [2]. Social network user sentiment analysis plays an important role in many aspects, such as political support, marketing and so on. Whether it is political election or marketing, we must pay attention to the emotional behavior of netizens. Therefore, the study of emotion is very important, which is the key factor affecting human behavior[3].

Therefore, taking Sina Weibo data as an example, this paper proposes a multi-level attention mechanism based sentiment trend prediction of microblog users. To predict the emotional trend, we need to collect the emotional characteristics of microblog users, and use multi-level attention mechanism to divide the emotional behavior level of users. In view of the characteristics of microblog users' emotion changes easily, the emotion transfer model of microblog users is established, and the direct fuzzy algorithm is used to stabilize the convergence value. Through the judgment of emotional similarity, the emotional trend of microblog users is predicted.

2. PREDICTION OF EMOTIONAL TREND OF MICROBLOG USERS

2.1 collection of emotional characteristics of microblog users based on multi-level attention mechanism

Using the text sentiment analysis tool rostea, we can calculate the positive and negative emotions of microblog content, but we can't calculate the user sentiment in a large number of microblog data [4]. In order to solve this problem, this research uses multithreading parallel computing rewriting software, which can calculate a large number of microblog data and store the results in MySQL database[5]. The results of the roster software showed that neutral affective values were distributed at [- 5, + 5], positive and negative affective values were at (+ 5, + ∞) and (- ∞, + 25), while [- 25 + 25] were at (- 25, + 25, + 25) and (- ∞) respectivel.

$$\begin{aligned}
&(25,+\infty) \leftrightarrow +3\\
&(15,25] \leftrightarrow +2\\
&(5,15] \leftrightarrow +1\\
&[-5,5] \leftrightarrow 0\\
&[-15,-5) \leftrightarrow -1\\
&(-25,-15) \leftrightarrow -2\\
&(-\infty,-25) \leftrightarrow -3
\end{aligned} \tag{1}$$

Affective computing is mainly used to analyze the emotion of text content, which is generally divided into three types: positive, negative and neutral. Microblog user groups are complex, and it is difficult to judge their emotional tendency. It is necessary to build an emotional database to determine the emotion expressed by the author, and score the emotion according to the test results of emotion mining of microblog sender [6]. The following is the structure of microblog user group emotional characteristics database as shown in **Figure 1**:

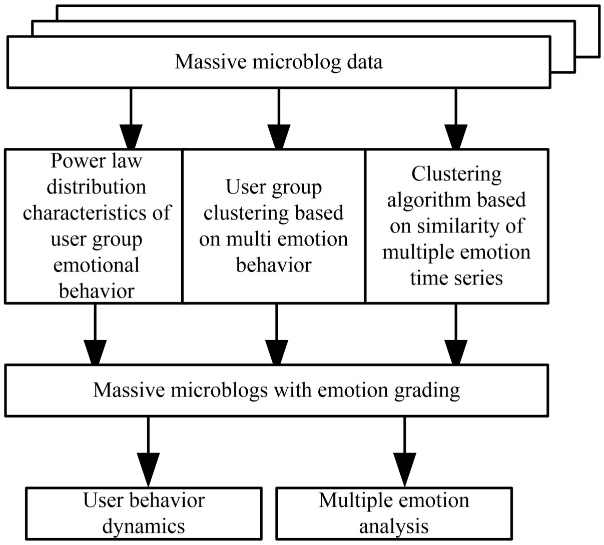

Figure 1 Structure of microblog emotional feature database

After dividing the emotional level of each microblog user in the database, the paper makes statistical analysis on the microblog users who published microblog from November to December 2019, a total of 12416 people. According to the emotional level (E), microblog texts are divided into three levels: the number of microblogs (n), the power index (α), the number of users (U), the number of microblogs (μ) and the number of variations (s). Based on this calculation, the specific calculation formula is as follows:

$$f_{\mathrm{ij}} = \frac{\alpha\mu - s}{2\sum \lim\limits_{5\to\infty} \mathrm{n}(E+U)} \quad (2)$$

When users express their emotions on microblogs, they are more likely to show calmness. For example, u, we can still see that the maximum value is μ when expressing neutral emotions, while this value increases slightly when expressing the two strongest emotions 3 and + 3. In this way, although the number of users expressing strong emotions will decrease, when these users express strong emotions, their participation in microblogging will be slightly improved[7]. Through the calculation of sample variance, it is found that the scale-free sample size has certain scale-free characteristics, and the scale-free characteristic of the sample is confirmed from one side, so as to divide the user's emotional behavior level, as shown in **Table 1**.

Table 1 Analysis of user's emotional behavior level

Similarity	Number of users 1	Number of users 2	Common users	Post quantity variance 1	Post quantity variance 2
1	19810	18614	17845	16844	15315
3	18218	17326	16180	15120	14014
5	17231	16524	15491	14602	13541
7	16541	15332	14295	13882	12546
9	15481	14340	13954	12361	11005
2,4,6,8	24508	23256	22052	21063	20151

According to table 1, the method in this paper is more suitable for text retrieval system, and its disadvantage is that it can not fully consider the relationship between texts, which leads to the independence of feature words and has certain relevance in practical application. In the process of feature extraction, the judgment function of each feature word should be fully considered.

2.2 Emotion transfer algorithm for microblog users

With the rapid development of social network, a large number of online interactive data have been generated, including users' thoughts, emotional tendencies and social interaction. Therefore, it is of great significance to study the relationship strength between overlapping and non overlapping groups and the overall emotional orientation between groups[8]. Through direct or indirect user interaction, researchers establish social networks and divide different groups. The multi-level attention mechanism is usually based on label passing, without considering the similarity between user nodes. Iterations are too random. The real large-scale social network may lead to the unreasonable distribution of the final community, and even the emergence of large-scale community, because there are too many iterations, it is difficult to make a choice. Based on this, the emotion transfer model of microblog users is constructed, as shown in **Figure 2**.

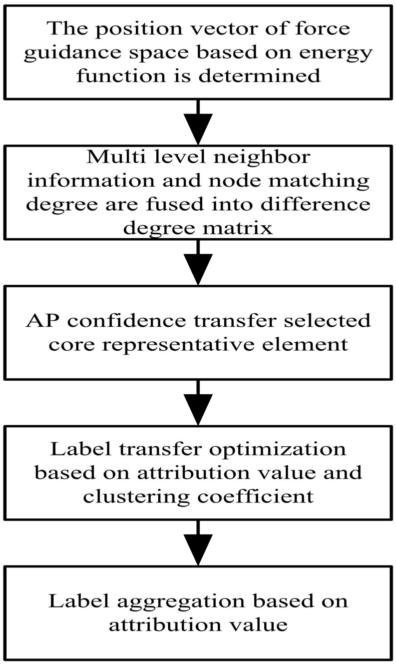

Figure 2 Emotion transfer model of microblog users

According to the principle of label transfer optimization, when the gain value of the emotional nucleon representation sequence of microblog users and the community division of the current adjacent points have obvious changes, in order to avoid local optimization, the attribute value results of emotional clustering of microblog users are taken as the basis of migration[9]. For the original community partition of any node, a unique initial label will be provided, which is regarded as the authorized community, and the node belonging value is introduced

$$g_i = A(i,i) + R(i,i) + \frac{f_i}{\sqrt{\sum_{j=1}^{N} f_j^2}} + \frac{G_i}{\sqrt{\sum_{j=1}^{N} G_j^2}} \quad (3)$$

$$G_i = \frac{w_{ij}^{'} / f_1 - \min_{j\in N}\left\{w_{ij}^{'} / f_1\right\}}{\max_{j\in N}\left\{w_{ij}^{'} / f_i\right\} - \min_{j\in N}\left\{w_{iji}^{'} / f_i\right\}} \quad (4)$$

In formulas (3) and (4), the A (i,j) represents the iterative convergence value of the attraction matrix, R (i,j) represents the iterative convergence value of the belonging matrix, and G represents the concentration degree of the nodes. Further, the weight of the network is redefined by the node's own attribution value.

$$w_{ij}^{'} = \begin{cases} w_{ij} & ,i \neq j \\ g_i & ,i = j \end{cases})(5)$$

On this basis, the weights gi of the original network are used to assign self weights to the nodes by the attribute values

$$w_i^{'} = g_i + \mathrm{c}_i \sum_{j\in\{x|\ x\neq l\}} w_{ij} \quad (6)$$

$$m^{'} = 2\sum_{i=1}^{N} w_i^{'} = 2m + \sum_{i=1}^{N} g_i \quad (7)$$

In the new network structure, ci represents the overall strength of the node structure and m represents the strength of the whole network. Taking the community division as the evaluation index, the new modular degree was calculated by formula (8).

$$\begin{aligned} Q_{nw} &= \frac{1}{m}\sum_i\sum_j\left(w_{ij}^{'} - \frac{w_i w_j}{m}\right)\cdot\sigma\left(C_i, C_j\right) \\ &= \frac{1}{m}\sum_1\sum_{j\neq i}\left(w_{ij} - \frac{w_i w_j}{m}\right)\cdot\sigma\left(C_i, C_j\right) + \frac{1}{m}\sum_{i=1}^{N}\left(g_i - \frac{w_i^{'2}}{m^{'}}\right) \\ &= \frac{1}{m}\sum_i\sum_{j\neq i}\left(w_{ij}^{'} - \frac{w_i w_j}{m}\right)\cdot\sigma\left(C_i, C_j\right) + Q_{sef} \end{aligned} \quad (8)$$

The clustering coefficient and attribute value are used to quantify the measurement coefficient of each node, and the corresponding sequence is obtained. Linear complexity is used to label convergence, which reduces the influence of random order on the stability of the algorithm. After the tags are converted, FRAP is grouped in the order of node measurement coefficients[10]. After modular optimization, the effect has reached the current optimal value, and the initial hierarchical structure of the community has been revealed. It can be divided into core cluster nodes, bridging nodes connecting different clusters and low membership noise nodes. Then, by integrating the historical series, the tag aggregation method is introduced, and the fusion coefficient considering the extreme value is introduced into the reconfigurable bridge nodes and noise points. In fact, it is a process to deal with the imbalance of importance differences. On this basis, the compactness and centrality information of the neighborhood of the node are extracted, and the combination coefficient and aggregation coefficient are extracted to measure the centrality and importance of the node. Based on this, the trend of user's emotion development is predicted.

2.3 Prediction of sentiment trend of microblog users

In recent years, scholars at home and abroad have done a lot of research on the posting behavior of netizens, which not only proves that microblog is a complex network, but also proves that its posting behavior obeys the power-law distribution, and establishes the corresponding posting model. But these studies can't distinguish the emotional level of each post[11]. In other words, this model can be used to simulate whether a user tweets at a certain time, but this model can not reflect the user's emotion on the microblog. The emotional behavior of users can be considered from three aspects: emotional communication between users and friends or familiar people; emotional stimulation of users themselves; emotional recovery of users themselves. However, this paper only makes a simple positive and negative segmentation of emotion, using abstract grid to simulate emotion, instead of using the real network. The specific steps are to analyze the emotional level of sina Weibo users when sending microblogs, and establish the emotional activity model of sina Weibo[12]. If the node sends microblogs at a certain point in time, et can be generated in two ways. When t + 1, the emotion level of node i is affected by the probability of its neighbor node at t time. That is, the emotional level of node j is affected by the probability of a neighbor node, while the node is affected by the probability of all neighboring nodes:

$$e_i^{(t+1)} = \begin{cases} e_j^t j \in N_i \wedge \delta_j^{(i)} = 1p \\ \left[\sum e_j^t\right] / \left(\sum \delta_j^{\prime i}\right) 1 - p_s \end{cases} \quad (9)$$

In formula (9), p represents whether neighbor node i has published microblog, e represents not publishing, j represents publishing; Ni represents all publishing nodes of neighbor node. In addition, node has an emotional release recovery process. If the node and its neighbors have no microblog or emotional fluctuation for a period of time, then the emotional level of node i returns to probability q, which further indicates the degree of convergence of the node's neighbors.For a given network Δ_i, each node can only be represented by the ratio of the adjacent nodes to each other.

$$c_1 = \frac{2\Delta_i}{k_i\left(k_i - 1\right)} \rho_t = \frac{f_i}{\sqrt{\sum_{j=1}^{N} f_j^2}} + \frac{g_i}{\sqrt{\sum_{j=1}^{N} g_j^2}} \quad (10)$$

In this case, Ki is the logarithm of neighborhood connectivity, and fi is the sum of neighborhood connectivity. The difference between the number of edge connections and the degree of adjacency can reflect the difference of importance between nodes to a certain extent. The surface layer is a measure of uncertainty, and the results of updating sequences with different initial nodes often lead to extremely unstable convergence values. To solve this problem, a direct fuzzy algorithm based on adjacency set potential is used to solve the problem of unstable convergence value[13]. The specific calculation is shown in formula (11).

$$H_v = -\sum_{t \in L\left(v, N_y\right)} P(l) \cdot \log P(S) \quad (11)$$

Here, P (s) is the node set with label set, and P (L) is the percentage of tag 1 in emotion set L (v, N). This scoring method can effectively reduce the overall mean square deviation of modular level test[14]. Clustering coefficient and information intensity partly reflect the importance and clustering strength of each node, and do not include the association information between two user nodes. Then, an association analysis method of social network nodes based on the concept of multi-user node strength is proposed[15]. However, this method only considers the relationship between direct nodes, but ignores the association between multi-user nodes, which results in the shortest path length of social network and poor correlation measurement between close nodes.

$$\sigma = \sigma_1\left(v_1, v_1\right) + m / ([n*(n-1)]/2) \cdot \left[\sigma_{21}\left(v_1, v_1\right) + \sigma_{22}\left(v_1, v_1\right)\right] \quad (12)$$

Through the above methods, the similarity v can be measured more truly, and the hidden correlation function n can be calculated. In addition, the node matching method is also a popular similarity measurement method. The algorithm uses the matching logarithm of nodes in two subgraphs to represent the similarity between nodes, and decomposes the information of network sub modules, and gives the measurement of node similarity. The dynamic model of social network emotion aims to construct the dynamic evolution system of social network emotion and reveal the internal mechanism of social network emotion occurrence, change and transmission. It is an interdisciplinary subject including psychology and computer science. And the research of these subjects does not exist independently, but promotes and depends on each other. Through the collection and summary of existing research, the overall framework of specific emotional trend prediction is shown in Figure 3.

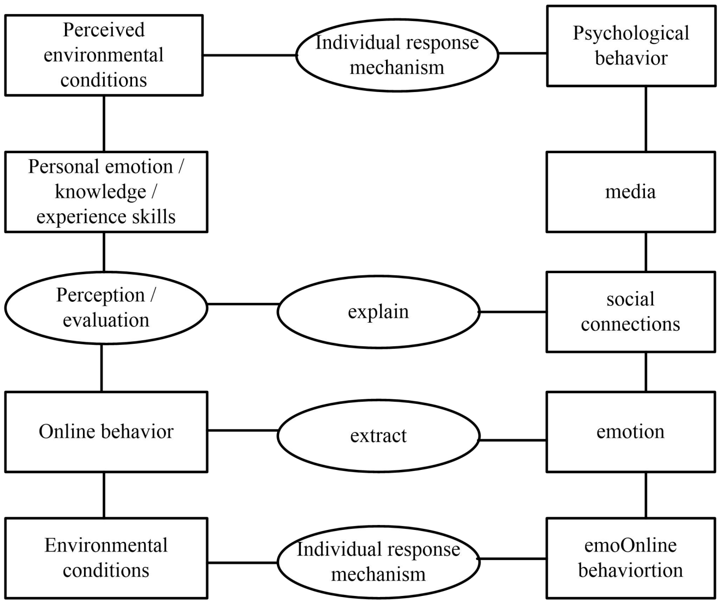

Figure 3 General framework of emotion trend prediction

The model in **Figure 3** combines two backgrounds: Psychology and computational science. Personal emotion is generated by the perception of environmental conditions, and the behavior triggered by individual reaction mechanism can include various emotional states. Finally, a person's emotion or behavior will reshape the original environment and promote a new round of emotion generation. The calculation of individual reaction mechanism can be regarded as a black box operation box, which can effectively analyze the emotional trend of microblog users and guide and control the network environment.

3. ANALYSIS OF EXPERIMENTAL RESULTS

In order to verify the effect of multi-level attention mechanism based sentiment trend prediction of microblog user groups, the data of microblog user groups from November to December 2019 were selected as historical data. Based on the simulated microblog social network, the MATLAB simulation model is established according to the above rules. The training set and test set are divided according to time. The division of experimental data is shown in **Table 2**.

Table 2 Experimental data of emotion prediction of microblog users

	Number of microblogs	Time
Training data	9678	November 21 - December 20
Test data	2652	December 21-22
User history	12416	

In the process of simulation, the time step of microblog network generation is set to 3000, and the number of user nodes is set to 300. The process of microblog network generation and model simulation are both random, and the model is simulated five times. In order to verify the accuracy of emotion prediction results and ensure the accuracy of prediction experiment answers, 2652 microblog text contents of test data were manually annotated, and 2168 data were obtained after de duplication.

According to the user posting behavior model of new microblog, considering the interaction between user emotion and adjacent points, the randomness of user emotion, and the ability of user emotion recovery, a user emotion model is established, and on this basis, a simulation comparative experiment is carried out. The comparison method is the method of reference [3]. The experimental results are shown in Fig. 4.

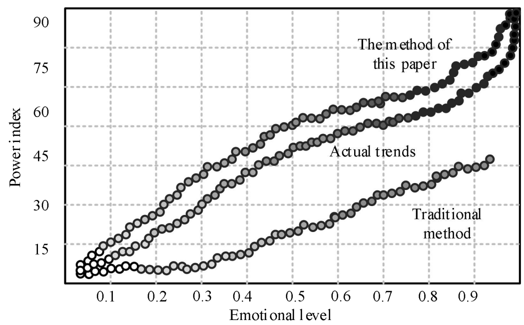

Figure 4 Prediction results of sentiment trend of microblog users

Based on the test results in Fig. 4, it can be seen that the experimental results are close to the actual microblog user group emotional change trend, and the research effect is obviously better, which fully meets the research requirements. The simulation results show that, in different emotional states, the number of Posts still follows the power-law distribution, the number of Posts increases when the emotional level tends to be stable, and the number of posts in positive emotional state is slightly larger than that in negative emotional state. The statistical experimental results are basically consistent in different emotional States, but not completely consistent. The focus of simulation is the change of the number of user posts with different emotion levels, so the power index itself does not affect the simulation results.

4. CONCLUSION

Due to the large number of microblog users and the different views and values expressed by users, their opinions and attitudes have universality and reference significance. In order to further improve the accuracy of micro blog users' emotional change trend and provide decision-making basis for social managers, this paper proposes a multi-level attention mechanism based micro blog user group emotional trend prediction. This paper constructs a micro blog user group emotional characteristics database from the perspective of social network user group emotional behavior, and divides micro blog users by multi-level attention mechanism Group microblog text emotional information, using node matching method to measure emotional similarity, complete the prediction of microblog user group sentiment trend. The simulation results show that the method proposed in this paper is close to the emotional trend of microblog users. However, due to the large population of microblog, there are still some differences in the research process. Therefore, the establishment of real emotional network is the next research direction.

REFERENCE

[1] Birjali M, Beni-Hssane A, Erritali M. Machine learning and semantic sentiment analysis based algorithms for suicide sentiment prediction in social networks[J]. Procedia Computer Science, 2017, 113: 65-72.

[2] Hong K , Liu G , Chen W , et al. Classification of the emotional stress and physical stress using signal magnification and canonical correlation analysis[J]. Pattern Recognition, 2017, 77(6):140-149.

[3] Li X, Rao Y, Xie H, et al. Bootstrapping social emotion classification with semantically rich hybrid neural networks[J]. IEEE Transactions on Affective Computing, 2017, 8(4): 428-442.

[4] Budiharto W, Meiliana M. Prediction and analysis of Indonesia Presidential election from Twitter using sentiment analysis[J]. Journal of Big data, 2018, 5(1): 51.

[5] Das S, Behera R K, Rath S K. Real-time sentiment analysis of twitter streaming data for stock prediction[J]. Procedia computer science, 2018, 132: 956-964.

[6] Ghosal D, Akhtar M S, Ekbal A, et al. Deep ensemble model with the fusion of character, word and lexicon level information for emotion and sentiment prediction[C]//International Conference on Neural Information Processing. Springer, Cham, 2018: 162-174.

[7] Agnès Ros-Morente, Cuenca E C , Guiu G F . Analysis of the Effects of two Gamified Emotional Education Software's in Emotional and Well-being Variables in Spanish Children and Adolescents[J]. International Journal of Emerging Technologies in Learning (iJET), 2018, 13(9):148.

[8] Pilato G , D'Avanzo, Ernesto. Data-driven Social Mood Analysis through the Conceptualization of Emotional Fingerprints[J]. Procedia Computer Science, 2018, 123(4):360-365.

[9] Ramadhani A M, Goo H S. Twitter sentiment analysis using deep learning methods[C]//2017 7th International annual engineering seminar (InAES). IEEE, 2017: 1-4.

[10] Ahmad A , Paul A , Din S , et al. Multilevel Data Processing Using Parallel Algorithms for Analyzing Big Data in High-Performance Computing[J]. International journal of parallel programming, 2018, 46(3):508-527.

[11] Sinha S, Saxena K, Joshi N. Sentiment Prediction of Facebook Status Updates of Youngsters[M]//Data,

Engineering and Applications. Springer, Singapore, 2019: 95-105.
[12] Naveh S , Bronstein J . Sense making in complex health situations: Virtual health communities as sources of information and emotional support[J]. Aslib Journal of Information Management, 2019, 71(6):789-805.
[13] Gutiérrez I, Gómez D, Castro J, et al. A new community detection algorithm based on fuzzy measures[C]//International Conference on Intelligent and Fuzzy Systems. Springer, Cham, 2019: 133-140.
[14] Nieuwazny J , Masui F , Ptaszynski M , et al. How religion and morality correlate in age of society 5.0: Statistical analysis of emotional and moral associations with Buddhist religious terms appearing on Japanese blogs[J]. Cognitive Systems Research, 2020, 59(Jan.):329-344.
[15] Hong K , Liu G , Chen W , et al. Classification of the emotional stress and physical stress using signal magnification and canonical correlation analysis[J]. Pattern Recognition, 2017, 77(10):140-149.