IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

SPECIAL SECTION ON INNOVATION AND APPLICATION OF INTELLIGENT PROCESSING OF DATA, INFORMATION AND KNOWLEDGE AS RESOURCES IN EDGE COMPUTING

Received November 29, 2019, accepted December 10, 2019, date of publication December 17, 2019, date of current version December 26, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2959831

User Multi-Modal Emotional Intelligence Analysis Method Based on Deep Learning in Social Network Big Data Environment

CHUNQIN ZHANG[1], **LICHUN XIE**[1], **YASEN AIZEZI**[2], **AND XIAOQING GU**[3]
[1]Zhejiang Industry Polytechnic College, Shaoxing 312000, China
[2]Department of Information Security Engineering, Xinjiang Police College, Ürümqi 830011, China
[3]School of Information Science and Engineering, Changzhou University, Changzhou 213164, China

Corresponding author: Yasen Aizezi (yasinjan@sina.com)

This work was supported in part by The National Natural Science Foundation of China (NSFC) under Grant 61762086, in part by the Zhejiang Natural Science Foundation Project under Grant y1080023, in part by the National Natural Science Foundation of China under Grant 61806026, and in part by the Natural Science Foundation of Jiangsu Province under Grant BK 20180956.

ABSTRACT In order to accurately analyze the emotional tendency of social media users' evaluation and better promote the work of emotional analysis and recommendation algorithm. This paper presents a new text emotion classification model, which integrates content features and user features, by representing the sentence, content features and user features of microblog as vector matrix and inputting them into the text emotion classification model which integrates content features and user features. Firstly, it analyzes the content information and user information related to the sentence emotion of the target microblog, and constructs the content characteristics and user characteristics respectively; Then, a text emotion classification model is constructed, which integrates content features and user features; Then, the method of feature level fusion and decision-making level fusion is designed for microblog emotion analysis of image and text fusion. Maxout neuron is also introduced to solve the problem of gradient dispersion in the training process and optimize the training process. The experimental results show that: Compared with other models, the accuracy of the proposed model is improved by more than 2.5%, and it is better than other models in recall rate and F value.

INDEX TERMS Multimodal affective analysis, deep learning, Maxout neuron, long and short term memory network (LSTM) model, intelligent analysis, feature fusion.

I. INTRODUCTION

With the development of social platforms such as microblog, more and more people like to express their opinions and feelings, spread content and exchange on the Internet. It has important theoretical value and practical significance to analyze the emotion of users publishing content on social platforms such as microblog. Consumers can provide reference for their purchase decision by understanding the attitude and evaluation of other users towards a product; by analyzing users' comments on their own goods and services, enterprises can grasp users' preferences, so as to improve products, improve efficiency and carry out personalized recommendation [1]–[3].

The associate editor coordinating the review of this manuscript and approving it for publication was Honghao Gao.

In natural language processing, emotional analysis has always been a key research topic. It analyzes, processes and summarizes the information with subjective emotional color, so as to judge the emotional tendency of users [4], [5]. As an important carrier of natural language, microblog emotional analysis has become one of the current research hotspots. Microblog is different from the general text. In a microblog message, in addition to the short text, it may also contain pictures, videos and other content, and microblog also has its own set of emoticons system, interaction and forwarding mechanism. As a social network platform, microblog users can understand the dynamic of users who are interested in it at any time through the attention mechanism. Users can interact in real time through comments, forwarding, likes and other behaviors, and each user in microblog has its own social relationship. Most of the existing microblog emotion analysis studies only focus on the content of the target microblog sentences, but ignore the emotional dependence of emoticons, pictures and microblog which also play an important role in microblog emotion analysis [6]. Different from the traditional text data, microblog messages are more difficult to analyze, and there are still many problems to be solved.

(1) Different from the traditional text, microblog messages contain special information such as emoticons, forwarding and comments. At the same time, each microblog user has its own unique attributes and emotional expression characteristics. The existing microblog emotional analysis methods do not fully consider these unique information and user information, which leads to the effect of emotional classification is not ideal.

(2) The existing microblog message is no longer a simple text form, but also contains pictures, videos and other information. Most of the research on microblog emotion analysis is based on microblog text, and some researchers consider the complementary role of text and text in microblog emotion analysis, but the constructed image emotion classification model has the problems of slow convergence or over fitting.

(3) As a social network platform, every user has his own social relationship, and the emotion between microblog is transitive and dependent. Most of the existing microblog emotion analysis methods simply assume that the emotions of different microblog are independent of each other, without considering the emotional dependence between microblog, so it is difficult to accurately determine the true feelings of users.

Therefore, in the research of microblog emotion analysis, in addition to the analysis of the target microblog sentences, the multi-modal data such as emoticons and pictures in microblog are also the focus of the research. In addition, the hidden user features, content features, interaction and forwarding features in microblog also play a key role in the emotional analysis of microblog. With the advent of the Internet era, the social network platform represented by microblog has become the main way for people to share and obtain information, resulting in a large number of data containing users' emotions. Emotional analysis of these microblog data is helpful to promote the research of public opinion analysis, personalized recommendation and emergency prevention. Therefore, microblog emotional analysis has important theoretical value. In recent years, more and more scholars have paid attention to it. Based on Sina microblog, this paper focuses on the most common image and text in microblog, focusing on the extraction and construction of feature information related to microblog sentence emotion, and using the method of deep learning to integrate the constructed features with the target microblog to improve the accuracy of microblog emotion classification. The main work completed includes:

The feature of this model is that it combines the advantages of deep learning model in abstracting abstract features with the idea of integrated learning and multi classifier decision-making. Maxout neuron is introduced on the basis of LSTM model. Although Maxout neuron increases the complexity of model in the process of model training, it is far less than the complexity of word vector transformation process, so it does not increase the complexity of the whole model in general. It solves the problem of gradient dispersion in the random gradient descent algorithm and optimizes the training process better. Finally, the feasibility and superiority of the model proposed in this paper are fully verified by the comparative test.

II. RELATED RESEARCH

A microblog message usually contains information such as pictures and texts, which are highly correlated in semantic space. It can provide complementary information for each other, give full play to the complementary role of different forms of data, and make more accurate judgment of emotional polarity. Reference [7] firstly extracts emotional features of text, voice and video, then splices the extracted emotional features, and finally inputs them into support vector machine (SVM) classifier for emotional classification. Reference [8] uses semantic dictionary, sentence structure rule and emotional tag as text emotional features, then uses the corresponding adjective noun pair in ANP as image emotional features, and finally uses parameters to control the fusion ratio of text and image for emotional classification of image text fusion. In reference [9], a cross media bag of words model (CBM) is proposed. The extracted text features and picture features are used to form a word bag model, and the logical regression, SVM and NB are used for comparative analysis. In reference [10], a new semi supervised emotion classification model CBOW-LR is proposed by extending CBOW model for distributed vector representation and emotion classification of text, and then the emotion features of pictures are extracted by using denoising automatic encoder. Finally, the emotion classification of picture and text fusion is carried out in an unsupervised and semi supervised way.

Deep learning method has been successfully applied in text and image single modal emotion analysis tasks. Using deep learning to do graph and text based cross modal emotion classification has become a new research field. In reference [11], CNN is used to extract text features at word level, phrase level and sentence level, and feature level fusion method is used to fuse text features at different levels with image features, so as to carry out emotional analysis of image and text twitter. In reference [12], we use two CNN to extract the features of text and image respectively, and then combine the features of full connection layer of text and drop connect layer of image. Finally, we use the method of logistic regression to fuse the features for emotional analysis. In reference [13], a cross modal consistent regression (CCR) scheme is proposed, which trains a regression model with deep text and picture features. In order to make use of the internal relevance between text and image, reference [14] uses the long and short term memory network (LSTM) model of tree structure to analyze the emotion of text and image, so as to better map the words and image regions.

Compared with other fusion models, this model has a better effect. In reference [15], attention mechanism is combined with CNN and SVM respectively, and image emotion classification model and text emotion classification model are constructed respectively. Finally, based on the proposed image emotion classification model and text emotion classification model, the effect of image and text emotion classification with different fusion methods is studied.

Microblog messages are different from traditional product reviews, news and other text content. Microblog messages contain special data such as emoticons, social relations and so on. Making full use of these special data in emotion analysis can improve the effect of emotion classification. Reference [16] proposes a microblog theme sentiment mining model which integrates emoticons and user's personality and emotion characteristics, which is used to judge the emotion between microblog sentences and microblog themes. Reference [17] crawls the interactive microblog related to the target microblog, the microblog with the same topic and the user's historical microblog as the training corpus, then extracts the words with higher term frequency–inverse document frequency (TF-IDF) value in the corpus as the context feature, and finally combines the context feature with the target microblog sentence as the input of convolution neural network model to classify Twitter's emotion. Reference [18] constructs a theme emotion model based on latent dirichlet allocation (LDA) and microblog user relationship, comprehensively considers the fact that microblog users are related to each other, and uses microblog user relationship to learn the emotional polarity of microblog. Reference [19] constructs semantic feature representation matrix for common microblog emoticons, and then uses multi-channel convolutional neural networks (CNN) for feature learning to achieve emotion classification.

Through the analysis of the research status at home and abroad, we can find that although the field of emotional analysis has achieved good results, the emotional analysis of microblog is still in the primary stage. Different from the traditional text data, microblog messages are more difficult to analyze, and there are still many problems to be solved. Based on this, a user multi-modal emotion analysis method using deep learning in social network big data environment is proposed.

III. CONTENT CHARACTERISTICS AND USER CHARACTERISTICS OF MICROBLOG

In microblog, in addition to microblog sentences, emoticons, user activity and the degree of microblog being forwarded and concerned may have a certain auxiliary effect on microblog emotional judgment. Therefore, we choose emoticons, the number of emoticons and the influence of microblog as the content features, and the user's activeness and the number of fans as the user features. For each type of feature, a series of keywords are selected for identification, and then the keywords are combined into feature sentences. The selection of content features and user features is based on the analysis of sina microblog dataset crawled on the Internet. Each microblog contains not only microblog sentences, but also microblog emoticons, microblog comments and forwarding numbers. Each microblog corresponds to a user. The user's relevant information includes the number of users' fans and the historical microblog published by the user in the last ten days from the target microblog (the microblog to judge the emotion). The historical microblog is used to calculate the user's activity.

A. CONTENT CHARACTERISTICS

Emoticons: Through the use of emoticons, microblog users can express their opinions and emotions quickly and intuitively. Emoticons have an important impact on the emotional expression of the text, and even can change the emotional polarity of the text. In order to describe the relationship between the number of emoticons and the emotional polarity of microblog, the cumulative distribution function (CDF) is introduced, and the formula is defined as formula (1).

$$F_x(x) = p(X \le x) \tag{1}$$

Through a large number of experiments on the data set, the statistical results show that the number of neutral microblog emoticons is almost less than 3, positive and negative microblog are more inclined to use more emoticons to express emotions, according to the number of specific emoticons, we can clearly distinguish the neutral and non neutral (positive or negative) microblog. Mark microblogs with less than or equal to 3 emoticons as ''less emoticons'', and those with more than 3 emoticons as ''more emoticons''.

Microblog influence: The influence of microblog is mainly reflected by the number of microblogs forwarding and comments. The more microblog forwarding and comments, the more attention microblog content receives. The calculation formula of microblog influence is as follows:

$$y_{i,j} = \sqrt[3]{MR_{i,j}} + \sqrt{MC_{i,j}} \tag{2}$$

B. USER CHARACTERISTICS

User activity: User activity has been studied in different forms of networks, such as communication networks and social networks. The user's activity is classified according to the frequency of user's participation in microblogging activities in unit time. First, the users of daily frequency ($f \le 3$) are divided into low active users, the users of daily frequency ($4 \le f \le 6$) are medium active users, and the others are high active users. Then, the daily frequency f was adjusted through comparative experiments. When the daily frequency of low activity degree was $f \le 2$, the daily frequency of medium activity degree was $3 \le f \le 5$, and the daily frequency of high activity degree was $f \ge 6$, the accuracy of the model was the highest.

Number of fans: To some extent, the number of fans reflects the influence of users. The users with more fans are generally celebrities or influential users. These users pay more attention to their public image and are more inclined to

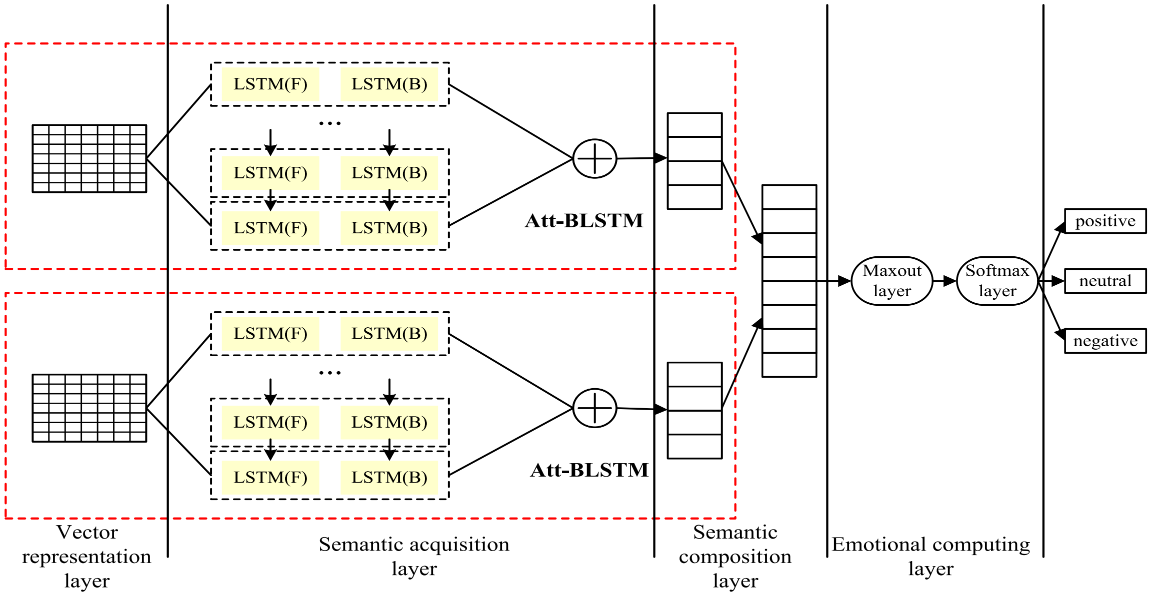

FIGURE 1. Structure of text emotion classification model.

publish active microblog or neutral microblog with positive energy. According to the research results, the number of followers of users who publish negative microblog is almost less than 150000. According to the specific number of followers, negative microblog and non negative microblog can be distinguished obviously. Users with more than or equal to 150000 fans in the data set are labeled as "multi fans", and users with less than 150000 fans are labeled as "few fans".

In the microblog data with emoticons, the number of emoticons varies in the range, but according to the statistical results, most microblog only have one or two emoticons. The influence of microblog is calculated according to the number of comments and forwards of microblog, and the data of microblog crawled is calculated. The change range of microblog influence changes between, most of which are in the range of. The data range of the number of users' fans is large, and the most users have tens of millions of fans. The statistical results of the number of emoticons, the influence of microblog and the number of users' fans are calculated based on the randomly crawled microblog data, which also has certain applicability for different time periods of microblog.

IV. A TEXT EMOTION CLASSIFICATION MODEL COMBINING CONTENT FEATURES AND USER FEATURES

A. MODEL COMPOSITION

By introducing attention mechanism into two-way LSTM, a text emotion classification model integrating content features and user features is constructed. The overall structure of the model is shown in Figure 1. The model is mainly composed of vector representation layer, semantic acquisition layer, semantic synthesis layer and emotion computing layer.

Vector representation layer: The vector representation layer is the input of the model, mainly including the word vector matrix R_F after feature sentence segmentation and the word vector matrix R_T after microblog sentence segmentation. Each vector matrix is $n \times d$, n is the number of words in the text, d is the dimension of word vector.

Semantic acquisition layer: This paper uses two-way LSTM and attention model to encode semantic information of text. In bidirectional LSTM, the output state at a certain time is connected by the output of forward LSTM and reverse LSTM at the current time, which is expressed as $h_t = \left[\overrightarrow{h_t}, \overleftarrow{h_t}\right]$. By multiplying the output h_t of the implicit state of each moment by their respective attention distribution weights and adding them $(a_1, \cdots, a_t)$, the calculation formula of the text semantic vector of the importance degree of each word in the text is as follows:

$$c = \sum_{t=1}^{T} a_t h_t \tag{3}$$

where T is the length of the text sequence, a_t is the attention weight of the t moment implicit state output h_t, indicating the importance of h_t, and a_t is calculated as follows:

$$f_{att}\left(h_t, s\right) = v \tanh\left(Wh_t + Uh_T + b\right) \tag{4}$$

where, v, W, U are training weight matrix, b are bias terms. Attention is paid to the calculation of the matching degree between the output h_t and the final output h_t of the hidden state at t moments by using the full connected network with hidden layer, then, Softmax was used to normalize, and the weight of attention distribution was a_t. Although the Maxout neuron increases the complexity of the model in the process of model training, it is far less than the complexity of the word vector transformation process, so it does not increase the complexity of the whole model in general.

Semantic composition layer: The main task of the semantic synthesis layer is to combine the semantic vector of the microblog sentence output by the model with the semantic vector of the feature sentence, and build the whole semantic vector.

Emotional computing layer: The main task of emotion computing layer is to build emotion classifier, calculate and output the final emotion classification results. The emotion probability distribution is calculated by softmax classifier, and the formula is as follows:

$$p_i(x) = \frac{\exp(x_i)}{\sum_{j=1}^{C} \exp(x_j)}, \quad i = 1, 2, \cdots C \tag{5}$$

C represents the number of emotion categories, and the calculation formula of loss function is:

$$loss = -\sum_{d \in T} \sum_{i=1}^{C} p_i^t(d) \log_2 \left(p_i^p(d)\right) \tag{6}$$

where, T is the training data set, d is a training sample, $p_i^t(d)$ is the true probability distribution of emotion classification, and $p_i^p(d)$ is the predicted probability distribution.

B. MAXOUT NEURONS

Although the traditional sigmoid function is widely used in the depth neural network, the problem of gradient dispersion often occurs in the training of random gradient descent. In the gradient descent algorithm, the fitting function is assumed to be

$$h_\theta(X) = \theta_0 + \theta_1 X \tag{7}$$

So the loss function is

$$J(\theta_0, \theta_1) = \sum_{i}^{n} (h_\theta(x_i) - y_i) \tag{8}$$

where: x_i is the ith element of sample feature X; y_i is the ith element of output Y. During the training process, the parameters are continuously updated until fitting

$$\theta_i = \theta_i - \alpha \frac{\partial J(\theta)}{\partial \theta_i} \tag{9}$$

where α is the learning rate.

The traditional batch gradient descent algorithm uses the whole sample set to iteratively calculate parameter θ. The complexity of each iteration is $O(mn)$. Where m is the number of samples and n is the number of elements in each sample feature X.When the sample m is large, in order to reduce the computational complexity, the random gradient descent algorithm is usually used. That is to say, every time a sample is read, the parameter θ is updated iteratively, so the complexity of one iteration is only $O(n)$, which greatly reduces the complexity of the algorithm. The random gradient descent

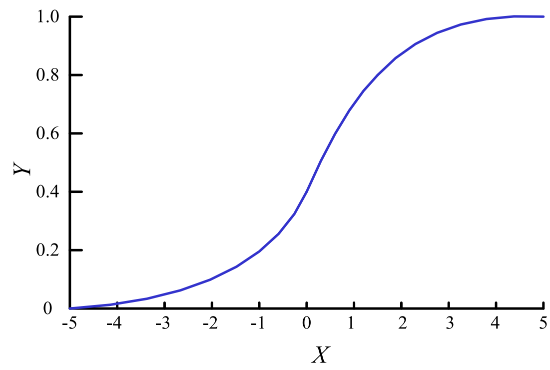

FIGURE 2. RBF neural network structure of Gaussian function.

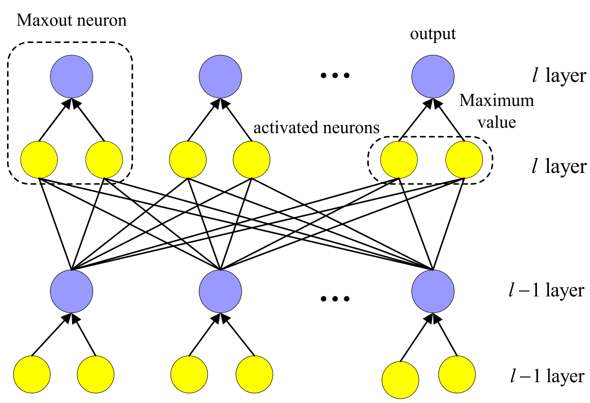

FIGURE 3. Structure of Maxout neurons.

algorithm is selected to train the model. The Sigmoid function is shown in Figure 2.

It can be seen from Fig. 2 that if Sigmoid is selected as the activation function, when the random gradient descent algorithm calculates the derivative, with the increase of network depth, the amplitude of gradient will sharply decrease, resulting in the slow change of the weights of the first several layers, so that they cannot get effective learning, that is, the so-called gradient dispersion problem [20], [21]. However, the emergence of Maxout neurons can solve this problem and increase the depth of the model. Its structure is shown in Figure 3.

It can be seen from Figure 3 that each Maxout neuron is composed of several different activated neurons, and its output is the maximum value [22], [23] of the activated neurons

$$h_j^i = \max_{j \in 1, \cdots, k} z_l^{ij} \tag{10}$$

where: h_j^i is the output of the ith Maxout cell in the lth layer, shown in blue; k is the number of activated neurons in the Maxout neuron; z_l^{ij} is the activation value of the jth activated neuron in the ith Maxout neuron in the lth layer, shown in yellow. The activated neuron is directly connected with the previous layer of neural unit through the weight matrix and bias vector, and is

$$z_l = W_{l-1}^T + b_l \tag{11}$$

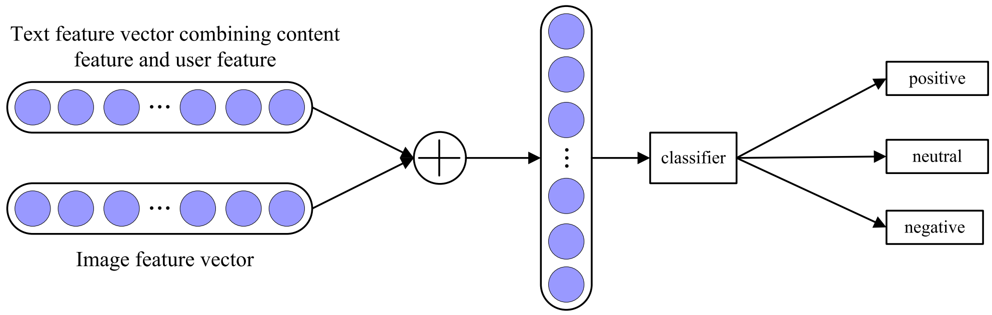

FIGURE 4. Comparison of classification of data sets of patients with coronary heart disease using different algorithms.

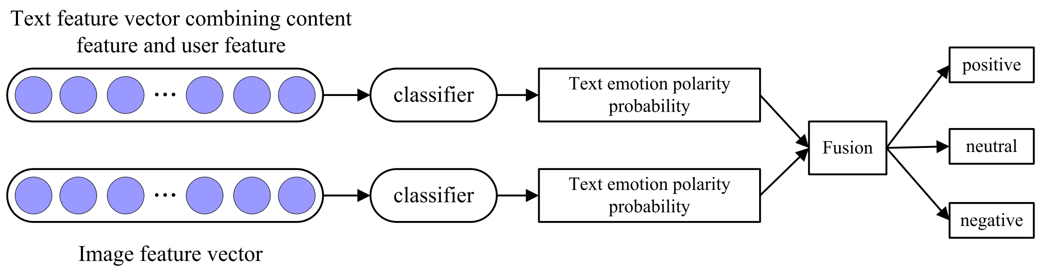

FIGURE 5. Composition of decision level fusion classification.

In the training process of random gradient descent algorithm, the gradient of Maxout neuron is

$$\frac{\partial h_l^i}{\partial z_l^{ij}} = \begin{cases} 1, & if\ z_l^{ij} > z_l^{is}, \quad \forall s \in 1, \cdots k; \\ 0, & otherwise \end{cases} \tag{12}$$

The above formula shows that the gradient of the output h_j^i of Maxout neuron to the activation unit z_l^{ij} with the maximum activation value is 1, and the rest is 0. It can be seen that Maxout neurons can solve the gradient dispersion problem by generating continuous gradient during the training process of random gradient descent [24]–[26]. Therefore, Maxout deep neural network has better training effect than traditional Sigmoid neural network.

C. FUSION CLASSIFICATION

Feature layer fusion classification: Feature layer fusion combines the feature vectors extracted from the text and image classification models respectively, and then inputs the combined feature vectors into the classifier for emotion polarity prediction. The fusion process is shown in Figure 4.

The text feature vector output from the full connection layer is recorded as V_t; then the image feature vector output from the full connection layer is extracted as V_i; finally, the vector connection V_t and V_i are used to connect and act as the feature vector V of the image and text:

$$V = V_t \oplus V_i \tag{13}$$

After obtaining the feature vector of image and text, the mapping from feature space to emotion space is completed by using Softmax classification algorithm.

Decision level fusion classification: Decision level fusion classification is to use the classifier to predict the emotional polarity probability of pictures and texts, then use the fusion formula to fuse, and finally get the emotional polarity of pictures and texts [27]–[29]. as shown in Figure 5.

In this paper, a weighted sum method is designed for decision level fusion. Each multimodal data sample in the dataset S, in the case of M modes and C emotion classifications, the probability set of emotion polarity is $\{P_{mc}(s), m = 1, 2, 3 \ldots M; c = 1, 2, 3 \ldots C\}$. Because only two modes of microblog image and text, positive, negative and neutral, M is 2, C is 3. There are three emotional probability values (positive, negative and neutral) for the final output of each mode, so the size of the emotional polarity probability set is 6 ($2 * 3$). The principle of weighted summation is that for each emotion classification c, c emotion probability values of M modes are weighted summation, and a weight coefficient is set before each mode, and the calculation formula of weighted summation is as follows:

$$p_c(s) = \sum_{m=1}^{M} \lambda_m p_{mc}(s) \tag{14}$$

$$\sum_{m=1}^{M} \lambda_m = 1 \tag{15}$$

λ_m is the weight coefficient of the mth mode. According to formula (14), a new probability set $\{p_c(s), c = 1, 2, 3 \ldots C\}$

is obtained. Finally, the emotion category corresponding to $p_c(s)$ with the largest probability value is the classification result after data fusion.

V. EXPERIMENT

A. DATA SET

Because multimodal emotional analysis is still in its infancy, and Sina microblog, Douban and other platforms impose various restrictions on the crawling data for the sake of privacy security, there are few standard data sets for experimental comparison in scientific research. The data set of image and text microblog used in the experiment is the image and text data set of Sina microblog crawled on the Internet.

First, we write the data of Sina microblog crawler crawling (microblog data includes microblog text, pictures, the number of microblog comments and forwards) and the user information related to microblog (the number of users' fans and the user's nearly ten day history microblog). Then, the data of microblog is preprocessed to remove advertisement and news. After preprocessing, 5 staff will mark manually. The marking rules are as follows: if there are three or more people in a microblog with the same tagging results, the emotional polarity of the microblog is the result of most taggers, and the ambiguous microblog will be deleted. Finally, 8745 positive microblog and 4685 neutral microblog and 2315 negative graphic microblog were obtained. After the data annotation, we extract the emoticons of each microblog, count the number of emoticons and calculate the influence of microblog to construct content features. Each microblog also has corresponding user features, which are mainly the number of users' fans and user activity. The user activity is calculated according to the number of microblog published by users in the last ten days.

B. EVALUATION INDEX AND PARAMETER

In order to verify the effectiveness of the text emotion classification method, image emotion classification method and image text fusion method proposed in this paper, accuracy, recall and F value are used as the evaluation indexes of classification effect [3], [30]–[32]. Before introducing the calculation methods of the accuracy rate, recall rate and F value of the three categories, the calculation methods of the two categories are introduced, and the calculation formula is as follows:

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{16}$$

$$recall = \frac{TP}{TP + FN} \tag{17}$$

$$F = \frac{Precision \times Recall \times 2}{Precision + Recall} \tag{18}$$

TP indicates that the positive class is predicted as a positive class, TN indicates that the negative class is predicted as a negative class, FN indicates that the positive class is predicted as a negative class, and FP indicates that the negative class is predicted as a positive class.

TABLE 1. Model comparison test.

Mode	Accuracy rate	Recall rate	F value
Reference [8]	0.853	0.843	0.854
Reference [12]	0.865	0.865	0.865
Reference [15]	0.876	0.867	0.871
Proposed model	0.896	0.895	0.895

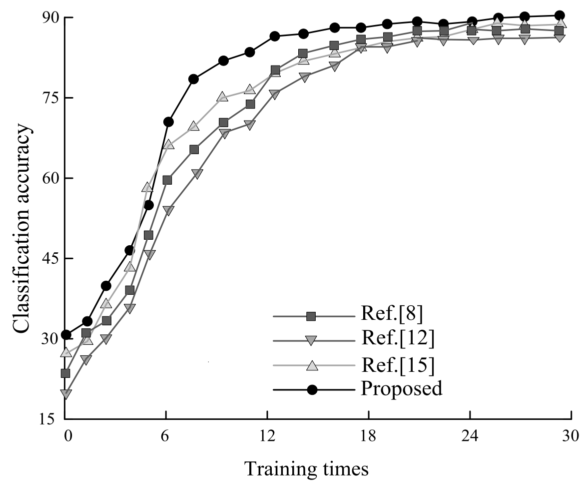

FIGURE 6. Classification accuracy of test sets in training process of each model.

The batch of the model is set to 100, and 100 samples are output each time for training. The optimization function is Adam [33], and the learning rate is set to 0.001. The training set and test set of the experiment are divided into 4:1. The experimental environment is Corei5 3.8GHz, Windows 10 64 bit system, 16g memory. Keras is used as the deep learning framework.

C. EXPERIMENTAL RESULTS AND ANAIYSIS

In order to verify the performance of the proposed model, the experimental results are shown in Table 1, compared with those in reference [8], reference [12] and reference [15].

From table 1, it can be seen that the proposed model has achieved the best results in all indicators. Compared with several comparative models, the average accuracy has increased by about 2.5%, which shows that the content features and user features constructed in this paper have a good indicative effect on emotion. The proposed model is superior to reference [8], reference [12] and reference [15] in accuracy, recall rate and F value, mainly because the proposed model introduces Maxout neuron on the basis of LSTM model to solve the gradient dispersion problem in the random gradient descent algorithm and optimize the training process better, so that the model has better classification effect.

Figure 6 shows the classification accuracy of all models in the training process. In the figure, the vertical coordinate is the classification accuracy, and the horizontal coordinate is the training times.

It can be seen from Figure 6 that with the increase of training times, the classification accuracy of each model gradually improves, and finally it stabilizes in a small fluctuation range. In reference [12], we use two CNN to extract the features of text and image respectively, and then splice the features of text full connection layer and image drop- connect layer. Finally, we use the method of logistic regression to fuse the features for emotional analysis. After 20 times of training, the accuracy rate stays at about 85.3%, which is lower than other models. Reference [8] uses semantic dictionary, sentence structure rule and emotional tag as text emotional features, and the accuracy stays at about 86.5%. Reference [15] combines attention mechanism with CNN and SVM respectively, constructs image emotional classification model and text emotional classification model, and achieves high classification accuracy, with the accuracy staying at about 87.6%. The proposed model introduces Maxout neurons on the basis of LSTM, and uses two independent LSTM forward and reverse splicing to extract features. It not only has content features and user features, but also introduces Maxout neurons to solve the problem of gradient dispersion in the training process and optimize the training process. The accuracy is about 90%, which is about 3% higher than the three comparison models. The comparison test fully verifies the feasibility and superiority of the model proposed in this paper.

VI. CONCLUSION

In natural language processing, emotional analysis has always been a key research topic. It analyzes, processes and summarizes the information with subjective emotional color, so as to judge the emotional tendency of users. As an important carrier of natural language, micro blog emotional analysis has become one of the current research hotspots. In the research of micro blog emotion analysis, in addition to the analysis of the target micro blog sentences, the multi-modal data such as emoticons and pictures in micro blog is also the focus of the research. In addition, the hidden user features, content features, interaction and forwarding features of microblog also play a key role in the emotional analysis of microblog.

This paper mainly aims at the problem that the effect of emotion classification is not ideal only through the target micro blog sentences, and improves the effect of emotion classification by integrating the target micro blog sentences with multi feature information.

This paper mainly aims at the problem that the effect of emotion classification is not ideal only through the target microblog sentences, and improves the effect of emotion classification by integrating the target microblog sentences with multi feature information. The sentences, content features and user features of microblog are represented as vector matrix, which is input into the text emotion classification model integrating content features and user features to realize the emotion classification of text. A user multi-modal emotion analysis method based on deep learning in the big data environment of social network is proposed.

However, due to the limitation of data set and other objective factors, there are still some areas to be further improved in this paper. Most of the existing microblog emotion analysis methods can not reflect the emotional expression characteristics of a single user when making emotional judgments, ignoring the user's personality and expression habits. Therefore, in the next step, we can try to build a separate emotional space for each user to classify user-oriented emotions.

REFERENCES

[1] H. El-Sayed, S. Sankar, M. Prasad, D. Puthal, A. Gupta, M. Mohanty, and C.-T. Lin, "Edge of things: The big picture on the integration of edge, IoT and the cloud in a distributed computing environment," *IEEE Access*, vol. 6, pp. 1706–1717, 2017.

[2] T. Horii, Y. Nagai, and M. Asada, "Modeling development of multimodal emotion perception guided by tactile dominance and perceptual improvement," *IEEE Trans. Cogn. Develop. Syst.*, vol. 10, no. 3, pp. 762–775, Sep. 2018.

[3] M. Huang, Q. Qian, and X. Zhu, "Encoding syntactic knowledge in neural networks for sentiment classification," *ACM Trans. Inf. Syst.*, vol. 35, no. 3, pp. 1–27, 2017.

[4] H. Gao, W. Huang, Y. Duan, X. Yang, and Q. Zou, "Research on cost-driven services composition in an uncertain environment," *J. Internet Technol.*, vol. 20, no. 3, pp. 755–769, 2019.

[5] K. Xia, H. Yin, P. Qian, Y. Jiang, and S. Wang, "Liver semantic segmentation algorithm based on improved deep adversarial networks in combination of weighted loss function on abdominal CT images," *IEEE Access*, vol. 7, pp. 96349–96358, 2019.

[6] K. Xia, H. Yin, and Y.-D. Zhang, "Deep semantic segmentation of kidney and space-occupying lesion area based on SCNN and ResNet models combined with SIFT-flow algorithm," *J. Med. Syst.*, vol. 43, no. 1, pp. 2:1–2:12, 2019.

[7] S. Poria *et al.*, "Towards an intelligent framework for multimodal affective data analysis," *Neural Netw.*, vol. 63, no. 6, pp. 104–116, 2014.

[8] D. Cao, R. Ji, and D. Lin, "A cross-media public sentiment analysis system for microblog," *Multimedia Syst.*, vol. 22, no. 4, pp. 479–486, 2016.

[9] M. Wang *et al.*, "Microblog sentiment analysis based on cross-media bag-of-words model," in *Proc. Int. Conf. Internet Multimedia Comput. Service*, 2014, pp. 76–80.

[10] C. Baecchi, T. Uricchio, and M. Bertini, "A multimodal feature learning approach for sentiment analysis of social network multimedia," *Multimedia Tools Appl.*, vol. 75, no. 5, pp. 2507–2525, 2016.

[11] G. Q. Cai and B. B. Xia, "Multimedia sentiment analysis based on convolutional neural network," *J. Comput. Appl.*, vol. 36, no. 2, pp. 428–431, 2016.

[12] Y. Yu, H. Lin, and J. Meng, "Visual and textual sentiment analysis of a microblog using deep convolutional neural networks," *Algorithms*, vol. 9, no. 2, p. 41, 2016.

[13] Q. You *et al.*, "Cross-modality consistent regression for joint visual-textual sentiment analysis of social multimedia," in *Proc. 9th ACM Int. Conf.*, 2016, pp. 13–22, doi: 10.1145/2835776.2835779.

[14] Q. You, L. Cao, and H. Jin, "Robust visual-textual sentiment analysis: When attention meets tree-structured recursive neural networks," in *Proc. 24th ACM Int. Conf. Multimedia*, 2016, pp. 1008–1017.

[15] Y. Chen and Z. Zhang, "IEEE 2018 13th IEEE conference on industrial electronics and applications (ICIEA)—Wuhan, China (2018.5.31–2018.6.2)," in *Proc. 13th IEEE Conf. Ind. Electron. Appl. (ICIEA)-Res. Text Sentiment Anal. Based CNNs SVM*, 2018, pp. 2731–2734.

[16] F.-L. Huang, S. Feng, D.-L. Wang, and G. Yu, "Mining topic sentiment in microblogging based on multi-feature fusion," *J. Comput.*, vol. 40, no. 4, pp. 872–888, 2017.

[17] Y. Ren, Y. Zhang, and M. Zhang, "Context-sensitive Twitter sentiment classification using neural network," in *Proc. 13th AAAI Conf. Artif. Intell.*, 2016, pp. 215–221.

[18] F. L. Huang, G. Yu, and J. L. Zhang, "Mining topic sentiment in microblogging based on micro-blogger social relation," *J. Softw.*, vol. 28, no. 3, pp. 694–707, 2017.

[19] Y.-X. He, S.-T. Sun, F.-F. Niu, and F. Li, "A deep learning model enhanced with emotion semantics for microblog sentiment analysis," *Chin. J. Comput.*, vol. 40, no. 4, pp. 773–790, Apr. 2017.
[20] K. Peddeti and S. Santhanam, "Dispersion curves for Lamb wave propagation in prestressed plates using a semi-analytical finite element analysis," *J. Acoust. Soc. Amer.*, vol. 143, no. 2, pp. 829–840, 2018.
[21] J. M. Gosselin, S. E. Dosso, and J. F. Cassidy, "A gradient-based model parametrization using Bernstein polynomials in Bayesian inversion of surface wave dispersion," *Geophys. J. Int.*, vol. 211, no. 1, pp. 528–540, 2017.
[22] H. Gao, W. Huang, X. Yang, Y. Duan, and Y. Yin, "Towards service selection for workflow reconfiguration: An interface-based computing," *Future Gener. Comput. Syst.*, vol. 87, pp. 298–311, 2018, doi: 10.1016/j.future.2018.04.064.
[23] L. Kostal and S. Shinomoto, "Efficient information transfer by Poisson neurons," *Math. Biosci. Eng.*, vol. 13, no. 3, pp. 509–520, 2017.
[24] Z. Jin, Y. Han, and Q. Zhu, "A sentiment analysis model with the combination of deep learning and ensemble learning," *J. Harbin Inst. Technol.*, vol. 50, no. 11, pp. 32–39, 2018.
[25] H. Gao, W. Huang, and X. Yang, "Applying probabilistic model checking to path planning in an intelligent transportation system using mobility trajectories and their statistical data," *Intell. Automat. Soft Comput. (Autosoft)*, vol. 25, no. 3, pp. 547–559, 2019.
[26] A. Zemmar, C.-C. Chen, and O. Weinmann, "Oligodendrocyte- and neuron-specific Nogo-a restrict dendritic branching and spine density in the adult mouse motor cortex," *Cerebral Cortex*, vol. 28, no. 6, pp. 1–9, 2017.
[27] B. Ando, S. Baglio, and C. O. Lombardo, "A multisensor data-fusion approach for ADL and fall classification," *IEEE Trans. Instrum. Meas.*, vol. 65, no. 9, pp. 1960–1967, Sep. 2016.
[28] W.-J. Lai, H.-K. Wang, H.-T. Liu, B. K. Park, S.-H. Shen, T.-P. Lin, H.-J. Chung, Y.-H. Huang, and Y.-H. Chang, "Cognitive MRI-TRUS fusion-targeted prostate biopsy according to PI-RADS classification in patients with prior negative systematic biopsy results," *J. Chin. Med. Assoc.*, vol. 79, no. 11, pp. 618–624, 2016.
[29] Q. Lu, X. Huang, J. Li, and L. Zhang, "A novel MRF-based multifeature fusion for classification of remote sensing images," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 4, pp. 515–519, Apr. 2016.
[30] W. He, Z. Zhang, and V. Akula, "Comparing consumer-produced product reviews across multiple Websites with sentiment classification," *J. Organizational Comput. Electron. Commerce*, vol. 28, no. 2, pp. 142–156, 2018.
[31] X. Chen, Y. Rao, and H. Xie, "Sentiment classification using negative and intensive sentiment supplement information," *Data Sci. Eng.*, vol. 4, no. 2, pp. 109–118, 2019.
[32] A. L. Bolaji, A. A. Ahmad, and P. B. Shola, "Training of neural network for pattern classification using fireworks algorithm," *Int. J. Syst. Assurance Eng. Manage.*, vol. 9, no. 1, pp. 1–8, 2018.
[33] B. Jou, T. Chen, and N. Pappas, "Visual affect around the world: A large-scale multilingual visual sentiment ontology," in *Proc. ACM Int. Conf. Multimedia ACM*, 2015, pp. 159–168.