An Emotion Analysis Model based on Fine-grained Emoji Attention Mechanism for Multi-modal We-Media

Chunxiao Fan, Siteng Chang, Yuexin Wu and Zheng Chen

October 14, 2020

An Emotion Analysis Model based on Fine-grained Emoji Attention Mechanism for Multi-modal We-Media

Chunxiao Fan, Siteng Chang, Yuexin Wu, and Zheng Chen

Beijing University of Posts and Telecommunications, Beijing 100876, China
cxfan@bupt.edu.cn

Abstract. As an isolated model, emoji plays a crucial role in emotion expression on We-Media platforms, which has always been the hot topic of NLP researches in the new environment. This paper analyses a relatively new method to such researches, namely, Emoji Attention Mechanism and analyses two of its main disadvantages: 1) it catches the literary meaning of the emojis' names only and discards the information illustrated from the emoji images, disturbing the final results of emotion classification; 2) it ignores the position of emojis in the text and it does not fit for the text emotion analysis that contains emojis in various clauses. Aimed to reduce the effect of the disadvantages mentioned above, this paper proposes the Semantic Vector Extraction based on Image(SVEI); taking the position of emojis in the text as a factor into consideration, this paper proposes an Emotion Analysis Model based on Fine-grained Emoji Attention Mechanism for Multimodal We-media, which enables the emotion of emojis to take effect in clauses more fine-grained. This paper adopts micro blog emotion analysis text published by NLPCC2014 as the data collection and designs the correspondent contrast experiments. This model raises the average F value and promotes the emotion identification accuracy by 3.18%.

Keywords: Emoji Attention Mechanism · Emotion Analysis · Semantic Vector Extraction.

1 Analysis on traditional emoji attention mechanism

Emotion analysis has experienced a long history of development: 1) traditional emotion analysis uses the emotion dictionary[1–4] at the beginning; 2) machine learning based on feature extraction[5–8]; 3) learning method based on the deep learning model[9–11]. Under the background of We-Media, the high accuracy of deep learning method, one of its advantages, can excavate the semantic information more effectively and correctly and is more suitable for the short text, a feature of We-Media platform. Thus, this paper uses the deep learning as the basic technology stack and builds up a deep learning model of multi-modal emotion analysis to fit for the diversified media information in We-Media platforms.

Each kind of source or form of information is called a kind of mode. Integrating various modes like emoji and text can deal with the emotion analysis on We-Media corpus more correctly. The strategies and methods of emoji emotion analysis mainly include following types[12–15]:

Type One, emojis are considered as a kind of note for emotion so that they can express users' emotion and opinions independently and then their emotion can be captured by setting up emoji-emotion dictionary. But such method has a disadvantage that it may cause ambiguity when the emoji exists in different contexts. And such dictionary contains a lot of noise, which will reduce the accuracy of emotion analysis model.

Type Two, emojis are integrated with emotion analysis model as a feature vector in addition.

Type Three, emoji and text are considered as two mutually independent information sources. Emotion scores expressed by emoji and text are calculated through different algorithm models respectively and then the final text emotion can be worked out by linear combination or simple summation of those scores.

Type Four, a relatively new method, it is a relatively representative structure model - emoji attention mechanism algorithm[16]. It measures the importance weight of the words in texts after which are combined with emojis, so as to indicate the emoji's effect on the text as well as on the meaning of the text. In this way, the expression of emotion polarity in the text can be changed.

Emotion expression is not a simple summation of text and emojis. The same emoji plays different roles in different texts. Emoji attention mechanism achieves a better analysis result on the corpus containing emojis but it is not completely applicable to the emotion analysis of We-Media, and remains suffering the following defects:

1) In the subsidiary task of emoji semantic vector extraction, it only learns the literary semantics of emojis' names, without considering the picture information carried by the emojis themselves. For example, when dealing with the emoji { 😘, kiss }, it simply take the unicode of the emoji "U+1F618" as the emoji's feature vector, or it will be regarded as UNK. However, We-Media users tend to pay more attention to the picture style presented by the emoji, rather than the name of it's when they are reading, so the traditional emoji attention mechanism undoubtedly misses the semantic information illustrated by the emoji pictures, then affecting the polarity of the final emotion classification.

2) It does not consider the position of emojis in the text, so it is not suitable for the text emotion analysis that contains emojis in various clauses. In this kind of text corpus, users often use emojis in a certain clause to enrich its emotion polarity. Thus, this emoji in the current clause have more obvious emotional effect than other clauses, especially in the clauses of such as shown in Table 1: when emojis in several clauses have emotional polarity conflict, the emoji in the former section of the sentence will weaken the emotional expression in the latter section. At this time, dividing the text range with each emoji has effect on is in greater demand. In the existing studies, it is roughly considered that all emojis have the same effect on all texts, taking the coarseness of the whole sentence as

the unit. Such work expands the influence range of emojis' weight and affects the accuracy of emotion analysis.

Based on the analysis above, this paper integrates the picture information of emojis and proposes the SVEI (Semantic Vector Extraction based on Image,SVEI). Besides, based on granularity, it advances the traditional emoji attention mechanism model and proposes the We-media Fine-grained Emoji Attention Mechanism (WFEA) which is suitable for the We-Media platforms.

Table 1. Corpus Example.

Text	Emotion
There are too many people having a cold or fever before the Spring Festival, and I cannot get away with it as well due to my poor health state {😭, tears }, I hope all of us can recover as soon as possible and enjoy our good time in this Spring Festival!{😏, titter}	Happy
It used to be a one minute journey, but now it takes 30 minutes to get home {😭, cry}. Streets and gardens are flooded with water but fortunately, it didn't flow into my house and doggies were standing at the higher place {👏, applause}. I wish it won't be worsen later.	Sad
Still cannot accept this result {🌚, moon face}. But what Italy brought to me has surpassed the happiness and touches of soccer itself, thank you Italy. God bless you{😊, smile}.	Like

2 Semantic Vector Extraction based on Image (SVEI)

Emojis in We-Media platform are often text information carried by pictures, which expresses users' abundant emotion. But when it comes to utilization, users tend to depend on their intuitive understanding. Thus, this paper adds image information as a factor and proposes the Semantic Vector Extraction based on Image (SVEI) that is integrated with image information.

This paper defines each piece of corpus is represented by $\{w_1, w_2, ..., w_T\}$ and $\{\{E_1, I_1\}, \{E_2, I_2\}, ..., \{E_K, I_K\}\}$. In this set, $\{w_1, w_2, ..., w_T\}$ is the result of words segmentation of sentences in the text. $\{\{E_1, I_1\}, \{E_2, I_2\}, ..., \{E_K, I_K\}\}$ is a set of emoji information included in corpus. E_i represents the name of the i-th emoji and I_i stands for the image information illustrated by the i-th emoji. SVEI algorithm procedure is shown by Fig. 1

In each emoji $\{E_i, I_i\}$, in the corpus, name vocabulary E_i are transformed into word vector e_i through Word Embed layer and emoji image I_i extracts feature vector i_i through convolution network layer. Then, e_i and i_i are sent to full connection layer for learning after being matched with each other. Finally, the encoding vector of this emoji v_i can be worked out.

$$\begin{aligned} v_i =& tanh(W_i[e_i; i_i] + b_i) \\ =& tanh(W_i[Embed(E_i); CNN(I_i)] + b_i) \end{aligned} \tag{1}$$

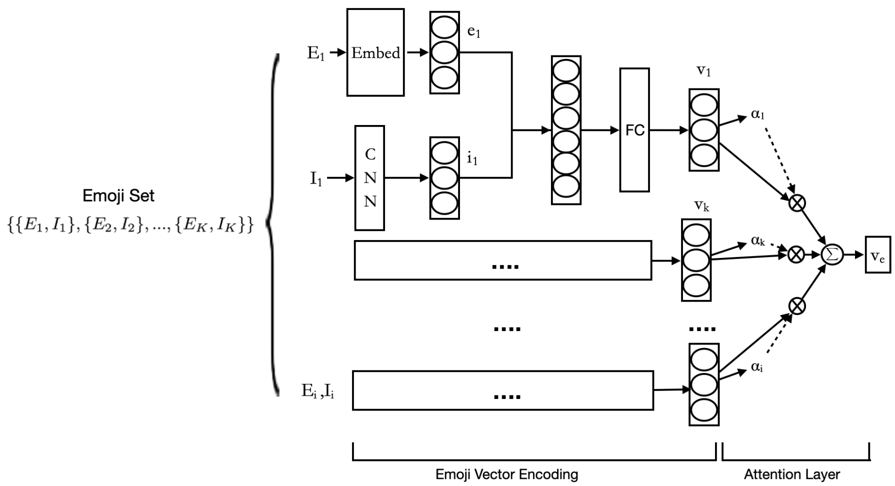

Fig. 1. The semantic vector extraction based on Image (SVEI) procedure.

Matrix W_i and constant b_i are the learning parameters of full connection layer.

The vector encoding matrix $[v_1; v_2; ...; v_K]$ learned from the emoji set. It is then sent to the attention layer to obtain the weight matrix $[\alpha_1; \alpha_2; ..., \alpha_K]$, and the weighted average is carried out to obtain the comprehensive semantic vector of the emoji set v_e.

$$\begin{cases} \alpha_i = softmax(V_w^T tanh(W_w v_i + b_w)) \\ v_e = \sum \alpha_i v_i \end{cases} \tag{2}$$

Matrix W_w, V_w and constant b_w are the network parameters of weight matrix in attention layer training.

3 We-media Fine-grained Emoji Attention (WFEA) Mechanism

According to users' habits of using emojis in We-Media platforms, they often add emojis in different places of the text to enrich the emotional expression for what they speak before. Emojis are weighted differently on the text, which affects the emotional effect. At the same time, the influence range of emojis varies in different places of the text. As what is described in the last chapter, when emojis in different places have emotional polarity conflict, the emoji in the former section of the text will weaken the emotional expression in the latter section. Therefore, this paper adds the emojis' position as a factor, proposes a granularity division and weight strategy based on the position of emoji, and limits the influence range of emojis' emotional weight to more granular clauses, so as to prevent the case that the influence range of emoji is so large that it disturbs the model effect.

With clauses as processing granularity, WFEA measures the emotional expression of each clause text combined with emojis at a smaller granularity, so as to more accurately capture the influence range of emojis, and thus be more suitable for the emotional analysis of We-Media platform.

WFEA model is shown by Fig. 2 and following is the certain algorithm procedure of this model.

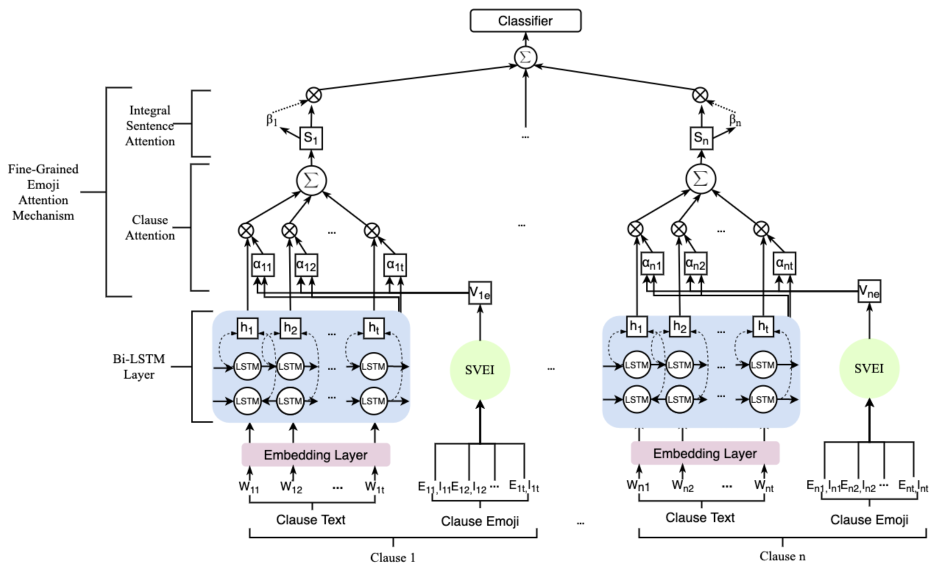

Fig. 2. The We-media Fine-grained Emoji Attention Model.

First, for each piece of corpus, emoji is considered as the separator of emoji influence range. Consider one corpus may contain several sentences, thecorpus is divided into $[\{[w_{11}, w_{12}, ..., w_{1T}]; [\{E_{11}, I_{11}\}, \{E_{12}, I_{12}\}, ..., \{E_{1K}, I_{1K}\}]\}; ...; \{[w_{n1}, w_{n2}, ..., w_{nT}], [\{E_{n1}, I_{n1}\}, \{E_{n2}, I_{n2}\}, ..., \{E_{nK}, I_{nK}\}]\}]$, where n stands for the number of sentences, K stands for the number of words in one sentence and T stands for the number of emoji in one sentence.

Second, for each sentence, extract the feature vector of emoji collection v_e through SVEI algorithm proposed in chapter 2 in this paper. Vector $\{v_{n1}, v_{n2}, ..., v_{nT}\}$ is the semantic vector acquired by the plain text $\{w_1, w_2, ..., w_T\}$ is learned through Word Embed Layer and LSTM layer. Meanwhile, in clause attention layer, words in the sentence are given different weights after they are combined with emojis through emoji attention mechanism.

$$\begin{cases} v_{emoji-ie} = SVEI([\{E_{i1}, I_{i1}\}, \{E_{i2}, I_{i2}\}, ..., \{E_{iK}, I_{iK}\}]) \\ v_{word-ij} = LSTM([w_{i1}, w_{i2}, ..., w_{iT}]) \\ \alpha_{ij} = softmax(score(v_{emoji-ie}, v_{word-ij})) \\ s_i = \sum_{j=1}^{T} \alpha_{ij} v_{ij} \end{cases} \tag{3}$$

α_{ij} represents the weight factor of the word w_{ij} after being combined with emoji semantic vector v_{ie} in the clause. s_i is the embedding vector of the i-th sentence in one corpus. Scoring function is applied to measure the influence on the integral meaning after text words are combined with emojis and remains to adopt the typical additive model statistical methods.

$$score(v_{emoji-ie}, v_{word-ij}) = V_s^T tanh(W_H v_{ij} + W_E v_{ne} + b_s) \tag{4}$$

Matrix W_H, W_E, V_s and constant b_s are learnable network parameters.

Third, the semantic vector matrix $s_1, s_2, ..., s_n$ is obtained after the clauses are learned and pieced together in step 2, and the weight matrix of each clause vector to the semantics of the whole sentence is trained through the attention mechanism model that accumulates the whole sentence level from bottom to top, represented as $\beta_1, \beta_2, ..., \beta_n$, and the comprehensive semantic vector S of the whole sentence is constructed by weighted average.

$$\begin{cases} \beta_i = softmax(V_N^T tanh(W_N v_i + b_N)) \\ S = \sum_{j=1}^{n} \beta_j s_j \end{cases} \tag{5}$$

Matrix V_N, W_N and constant b_N are learnable network parameters.

Finally, the emotion classification layer is carried out based on the integral semantic vector S.

4 Experiment Validation

4.1 Data Set

Based on the published Chinese data set NLPCC2014 and the emoji picture set collected by ourselves, this paper carries the following experiments.

NLPCC2014, the data set from Sina Weibo for Chinese micro blog emotion analysis and evaluation task, contains 14,000 noted training micro blogs and 6,000 test micro blogs. Each micro blog is noted according to sentence level respectively, involving 8 kinds of fine-grained emotion categories of anger, disgust, fear, happy, like, sad, surprised, and no emotion.

In this paper, each marked sentence is extracted, and the website address, user and topic tag in the data set are replaced. Jieba Chinese word segmenter is applied for word segmentation tasks.

The number of training corpus sample is 45421, different emojis 372, the total number of emojis appeared are 5313. The number of testing corpus sample is 15693, different emojis 316, the total number of emojis appeared are 4091, and the numbers of other emoji types are shown in Table 2.

Table 2. Corpus Statistics.

Emotion	Training Set	Testing Set
angry	1661	243
disgust	3361	675
fear	298	64
happy	2725	639
like	4333	1625
sad	2485	317
surprise	827	259
no emotion	29731	11871

Since the original data set does not contain the image information of emojis, this paper extracts the picture set of emojis on this platform through the source platform of the data set, namely the open public API of Sina Weibo, and constructs the additional picture data set of emojis, whose statistical data are shown in Table 3.

Table 3. Additional Emoji Dataset Statistics.

Total Sample	Sample Format	Figure Size	Color Space
1068	.png	36*36	RGB

4.2 Experiment Design

After pre-processing and word division of the corpus, the Glove model is used for pre-training. The represent dimensions of word vectors are set 300, thus obtaining the corpus word vectors and semantic information of the emojis. LSTM is adopted as neural network for learning, and the vector of each hidden node is 128 dimension. During the training, Adam algorithm is used to optimize the parameters, and the number of training rounds is 20.

Since this experiment divides the emotion granularity into 8 polarities, it belongs to the emotion multi-classification task. Taking the test set described in chapter 3.3 as the test expectation, this paper evaluates the experimental effect of the emotion classification model through the three indexes shown by the model in the test corpus: accuracy, Macro F1 value and Micro F1 value.

This experiment mainly tested the influence of different emoji analysis strategies on emotion analysis results, so the four baselines were compared:

1) LSTM-text. Plain text is input into the LSTM emotion analysis model. Emojis in corpus are filtered out and plain text is used as the input of the model. The semantic expression obtained by LSTM is directly used as the feature vector of the classification network. The optimization method, training method and parameter setting are the same as experiment settings.

2) LSTM-emoji. Emojis in corpus are retained, and the idea of combining the emotions of text and emojis is used to integrate plain text and emojis into the LSTM emotion analysis model as two features respectively. The model output is directly used as the feature vector of the classification judgment model. The optimization method, training method and parameter setting are the same as experiment settings.

3) LSTM-emoji-Attention. The traditional emoji attention mechanism is used and the text and emojis are retained. The average value of the emoji vector is taken as the attention feature of the text semantics of the LSTM model, and the comprehensive semantic expression obtained under the attention mechanism is taken as the feature vector of the classification network. The optimization method, training method and parameter setting were the same as experiment settings.

4) LSTM-WFEA. An emotion analysis model is constructed according to the We-media Fine-grained Emoji Attention Mechanism WFEA and LSTM. The extracted semantic vector of the emojis is taken as the attention feature, and the semantic vector expression of each clause based on LSTM and attention mechanism is obtained. Then, the comprehensive expression of all clauses is obtained through the whole sentence class attention mechanism learning, which is regarded as the feature vector of classification network. The optimization method, training method and parameter setting are the same as experiment settings.

4.3 Experiment Results and Analysis

From the perspective of the performance of the model in full picture, Table 4 shows the evaluation indexes of each model on the test set in this experiment.

Table 4. Experimental results of emotion classification.

Model	Acc(%)	Marco F1(%)	Micro F1(%)
LSTM-text	85.78	75.10	86.88
LSTM-emoji	86.19	75.88	87.31
LSTM-emoji-Attention	87.25	79.31	87.71
LSTM-WEFA	**88.30**	**79.59**	**88.05**

Comparing and analyzing the experiment data in Table 4, this paper finds that:

a) The model considering emojis and text multi-model information has higher accuracy rate of emotion classification and F1 index than the single modal model LSTM-text considering text information merely, which verifies the influence of emojis on emotion and the importance of emojis' feature introduction in emotion classification task.

b) In the case that emoji modal information is also considered, the traditional emoji attention mechanism model LSTM-emoji-Attention has a higher accuracy Acc by 1.06% than the non-emoji attention mechanism model LSTM-emoji. It is because there are many ironic, funny emojis in the corpus that cause changes in the emotional polarity of the text. The emoji attention mechanism can obtain more effective semantic vectors, which also verifies the effectiveness and feasibility of the traditional emoji attention mechanism.

c) In this paper, the We-media Fine-grained Emoji Attention Mechanism WFEA has higher accuracy Acc by one more 1.05% than traditional emoticons attention mechanism, and also has the Macro F1 and Micro F1 improved. This is because a wide range of emojis utilization causes various clauses with various emojis in the identified corpus. WFEA divides the text into the level of clauses as the granularity, which can more accurately define weights influence range of emojis.

From the specific performance of the model in each category of samples, **Appendix I** shows each indicator of each experimental model on different emotional polarity samples. As can be seen from the table, the We-media Fine-grained Emoji Attention Mechanism WFEA proposed in this paper has been improved in all indicators of the polar emotion samples, such as "anger, disgust, surprise, like", while the identification accuracy P of neutral and emotionless text is low and the recall rate R is high. This is because emojis often have a great influence on the emotional polarity of text, but have little influence on neutral text, which also supports the mechanism of emojis' positive effect on discriminate emotional polarity of text.

5 Conclusion

Taking image information of emoji into account can improve the ability of expression for semantic vectors. accurately extract the emotion feature vector contained by emojis. In this way, the semantic expression of each clause combined with emoticons can be obtained more effectively, so as to improve the overall accuracy of emotion classification.

Through the experiment results and comparative analysis above, in emotion classification task, the effectiveness and superiority of the We-media Fine-grained Emoji Attention Mechanism WFEA proposed in this paper are verified. Compared with the traditional emoji attention mechanism, the researches in this paper more effectively discovers the semantic weight and emotional influence between text and emojis, which improves the identification accuracy and effect.

References

1. Wu J, Lu K, Su S, et al. Chinese Micro-Blog Sentiment Analysis Based on Multiple Sentiment Dictionaries and Semantic Rule Sets[J]. IEEE Access, 2019, 7: 183924-183939.
2. Taboada M , Brooke J , Tofiloski M , et al. Lexicon-Based Methods for Sentiment Analysis[J]. Computtional Linguistics, 2011, 37(2):267-307.
3. Cheng-Gong Z , Pei-Yu L , Zhen-Fang Z , et al. A sentiment analysis method based on a polarity lexicon[J]. Journal of Shandong University(Natural ence), 2012, 47(3):47-50.
4. Wang Ke, Xia Rui. A survey on automatical construction methods of sentiment lexicons. Acta Automatica Sinica, 2016, 42(4): 495−511.
5. Sebastiani, Fabrizio. Machine learning in automated text categorization[J]. ACM Computing Surveys, 2002, 34(1):1-47.
6. Pang B , Lee L , Vaithyanathan S . Thumbs up? Sentiment Classification using Machine Learning Techniques[J]. Empirical Methods in Natural Language Processing, 2002:79-86.
7. Wawre S V, Deshmukh S N. Sentiment classification using machine learning techniques[J]. International Journal of Science and Research (IJSR), 2016, 5(4): 819-821.
8. Pang B, Lee L. A sentimental education: Sentiment analysis using subjectivity sum-marization based on minimum cuts[C]//Proceedings of the 42nd annual meeting on Association for Computational Linguistics. Association for Computational Linguistics, 2004: 271.
9. He Y X, Sun S T, Niu F F, et al. A deep learning model enhanced with emotion seman-tics for micro blog sentiment analysis[J]. Chin. J. Comput., 2017, 40(4): 773-790.
10. Zhang L, Wang S, Liu B. Deep learning for sentiment analysis: A survey[J]. Wiley In-terdisciplinary Reviews: Data Mining and Knowledge Discovery, 2018, 8(4): e1253.
11. Tang D, Qin B, Liu T. Deep learning for sentiment analysis: successful approaches and future challenges[J]. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 2015, 5(6): 292-303.
12. Kimura M, Katsurai M. Automatic construction of an emoji sentiment lexicon[C]//Proceedings of the 2017 ieee/acm international conference on advances in so-cial networks analysis and mining 2017. 2017: 1033-1036.
13. Wolny W. Sentiment analysis of Twitter data using emojis and emoji ideograms[J]. Studia Ekonomiczne, 2016, 296: 163-171.
14. LeCompte T, Chen J. Sentiment analysis of tweets including emoji data[C]//2017 In-ternational Conference on Computational Science and Computational Intelligence (CSCI). IEEE, 2017: 793-798.
15. Hogenboom A, Bal D, Frasincar F, et al. Exploiting emojis in sentiment analysis [C]// Proc of the 28th Annual ACM Symposium on Applied Computing. New York: ACM Press, 2013: 703-710.
16. Hao T , Shuwen D , Tao Q , et al. Emoji-attentional neural network for microblog sen-timent analysis[J]. Application Research of Computers, 2019.

A Appendix I

Table 5. Experiment results on different emotion polarity samples..

Experiment Model		Neutral	Disgust	Fear	Sad	Angry	Surprise	Happy	Like
LSTM-text	Acc(%)	90.32	96.37	99.24	97.62	98.06	98.71	97.59	95.84
	P(%)	**93.07**	75.08	45.23	80.25	73.83	62.62	80.59	76.93
	R(%)	92.09	74.47	75.08	74.62	73.07	72.75	78.09	**80.42**
	F1(%)	92.58	75.77	56.46	**77.33**	73.44	67.30	79.32	78.64
LSTM-emoji	Acc(%)	90.59	96.83	99.34	97.42	97.95	98.69	97.58	96.22
	P(%)	92.89	81.03	49.89	76.58	71.49	61.48	79.23	80.68
	R(%)	92.75	**74.89**	76.77	75.64	**73.43**	74.57	**80.03**	79.23
	F1(%)	92.82	77.84	60.84	76.11	72.45	67.40	**79.63**	79.95
LSTM-emoji-Attention	Acc(%)	**90.67**	97.02	**99.73**	**97.66**	98.22	99.04	97.68	95.81
	P(%)	91.68	84.12	**81.65**	**81.76**	77.63	72.93	83.31	76.99
	R(%)	94.33	73.75	**76.43**	73.24	72.10	**75.06**	76.14	79.81
	F1(%)	**92.99**	78.60	**78.96**	77.27	74.77	73.98	79.56	78.37
LSTM-WFEA	Acc(%)	90.54	**97.07**	**99.73**	97.46	**98.36**	**99.13**	**97.70**	**96.11**
	P(%)	90.97	**85.48**	81.07	77.29	**81.81**	**77.94**	**83.89**	**81.10**
	R(%)	**94.99**	72.89	**76.43**	**75.56**	71.14	72.63	75.77	77.12
	F1(%)	92.94	**78.69**	78.68	76.42	**76.10**	**75.19**	79.62	**79.06**