A multi-modal and multi-scale emotion-enhanced inference model based on fuzzy recognition

Yan Yu^{1,2} · Dong Qiu³ · Ruiteng Yan¹

Received: 6 September 2021 / Accepted: 26 October 2021 / Published online: 13 November 2021
© The Author(s) 2021

Abstract

Only the label corresponding to the maximum value of the fully connected layer is used as the output category when a neural network performs classification tasks. When the maximum value of the fully connected layer is close to the sub-maximum value, the classification obtained by considering only the maximum value and ignoring the sub-maximum value is not completely accurate. To reduce the noise and improve classification accuracy, combining the principles of fuzzy reasoning, this paper integrates all the output results of the fully connected layer with the emotional tendency of the text based on the dictionary to establish a multi-modal fuzzy recognition emotion enhancement model. The provided model considers the enhancement effect of negative words, degree adverbs, exclamation marks, and question marks based on the smallest subtree on the emotion of emotional words, and defines the global emotional membership function of emojis based on the corpus. Through comparing the results of CNN, LSTM, BiLSTM and GRU on Weibo and Douyin, it is shown that the provided model can effectively improve the text emotion recognition when the neural network output result is not clear, especially for long texts.

Keywords Deep learning · Sentiment analysis · Emoji · Fuzzy recognition · Lexicon-based sentiment enhancement

Introduction

People express their emotions in many ways, such as text [1,2], voice [3], intonation [4], facial expressions [5,6] and the multimodal combination [7]. With the popularity of online media, emojis have become one of the main ways people express their emotions intuitively. Existing work usually uses contextual emotions to predict or recommend suitable

emoticons, while ignoring the enhancement effect of emoji on contextual emotion.

Human emotions are complex, expressing multiple emotions at the same time [8]. In Chinese expression, sentences composed of the same words can express different emotional tendencies. The objects modified by negative words in different positions are also different. Emojis as a vehicle to intentionally convey affective states, can express the sentiment and emotion in texts. Take the following sentence as an example.

S_1 :
 S_1 : 就 差 长 江 索 道 没 有 坐 过
了 , 算 遗 憾 不 ?

Only from the textual analysis, different people have different emotional tendencies for the same text. People with positive emotions will feel that they have visited all the other scenic spots except the Yangtze River Cableway, which is no regret; negative people will think that it is regrettable that the Yangtze River Cableway did not go. If there are emojis in S_1 , we can get the most accurate emotional tendency based on the emojis, as shown in Table 1. A very happy emoticon in S_2 expresses positive emotional tendencies; while in S_3 a crying

✉ Dong Qiu
qiudong@cqupt.edu.cn

Yan Yu
yanyu2002034@163.com

Ruiteng Yan
s180231039@stu.cqupt.edu.cn

¹ School of Computer Science and Technology, Chongqing University of Posts and Telecommunications, Chongqing, China

² Chongqing College of Mobile Communication, Chongqing, China

³ College of Science, Chongqing University of Posts and Telecommunications, Chongqing, China

Table 1 Comparing to the sentiment influenced by emojis

No.	document	sentiment
S ₁	就差长江索道没有坐过了，算遗憾不？	positive
	(Only the Yangtze River Cableway has not been taken, is it a pity?)	negative
S ₂	就差长江索道没有坐过了，算遗憾不？😄	positive
	(Only the Yangtze River Cableway has not been taken, is it a pity?)	
S ₃	就差长江索道没有坐过了，算遗憾不？😄	negative
	(It's still a pity that you haven't taken the Yangtze River Cableway?)

Fig. 1 Overall architecture of the provided model

emoticon expresses negative emotional tendencies. Integrating the emotion of emoji into the sentiment analysis of the text can more accurately express the sentiment tendency of the text.

Some works take negative words into account when analyzing the sentiment of a text. Wu et al. [9] discussed the change of negative words to emotional polarity of emotional words. If an adjoint word of the emotional word is negative, the emotional polarity of the emotional word will be the opposite. Yan et al. discussed the influence of negative words and emojis on the polarity of emotions, ignoring whether negative words and emojis modify emotion words [10]. On these basis, this paper proposes a text emotion enhancement model based on the smallest subtree, discusses the emotion enhancement effects of negative words and degree adverbs on emotion words, and considers the enhancement of text emotion by question marks and exclamation points. Since fuzzy reasoning can not only deal with the fuzzy expression

of language [11,12], but also optimize the model without coupling relationship between components [13–15], we combine the emotional word dictionary, emoticon dictionary and deep learning model to establish an fuzzy recognition emotion enhancement reasoning model that optimizes the output results of the deep neural networks. The overall architecture of sentiment analysis model fused by deep neural networks, emojis and lexicon-based sentiment enhancement fuzzy reasoning is listed in Fig. 1.

The advantages of this article are as follows. (1) Construct a membership function of the emotion of emojis, and concatenate the emoji2vec with the word2vec of the input document. (2) Establish a lexicon-based fuzzy inference emotional enhancement model which consists of the emotional enhancement of negative words and degree adverbs on their modified emotional words and the enhancement of question marks and exclamation marks on text emotions. (3) Present a fuzzy classification membership function of the

fully connected layer of neural networks combined with the membership. (4) The fuzzy classification membership function of the fully connected layer of the neural network, the emotional membership of emojis and the dictionary-based fuzzy inference model are combined to establish a multi-modal and multi-scale emotion-enhanced inference model based on fuzzy recognition.

The rest of this paper is organized as follows. The next section summarizes some related literature on multimodal sentiment analysis and emoji in sentiment analysis. “3” explains our proposed model on sentiment enhancement of the output of deep neural networks by lexicon-based fuzzy reasoning model in detail. “4” shows the experiment processing. “5” demonstrates the experimental results and discussions. In the last section, some conclusions are drawn.

Related work

Multimodal sentiment analysis

Sentiment analysis has widespread commercial applications and practical values in various domains, such as decision analysis [16] and topic detection [17]. There are two main models to analyze the sentiment of a text: lexicon-based method [2] and deep learning-based method [9,18,19]. A labelled corpus with moral foundation scores was introduced to illustrate how morality-related information be inferred from text by shallow and deep learning models [2]. Residual memory networks can be used to comprehend multimodal human sentiment [20]. Huang et al. constructed a sentiment convolutional neural network to analyze the sentence sentiments based both on contextual and sentiment information of sentiment words [21].

The sentiment and emotion analysis of multi-modal dialogue such as video, audio, image attracts more and more attention [7,22–24]. To improve the performance of the classifier, the image-text pair in the existing unimodal data is usually used as a multi-modal comprehensive data set [25]. To obtain the true sentiment, significant leaps responsible for its current relevance by pointing out the shortcomings and under-explored was analyzed [26]. Considering both textual information and sentiment diffusion patterns, sentiment diffusion patterns were utilized to improve sentiment analysis in Twitter [27]. To rich correlations among social images, correlations among multiple social images were incorporated to conduct visual sentiment analysis [28].

Emoji in sentiment analysis

With the continuous upgrading of smart devices, the number and types of emojis are constantly upgrading, and people are more and more like to use emojis to express emotions.

In recent years, the research of emoji has attracted more and more attention. Emojis are studied in some computer-mediated communications [29], such as twitter [30,31] and posts [32–34].

Some scholars have studied the semantic information of emoji from the semantic aspect [35]. A model for emoji similarity was explored to enable moving forward to novel emoji keyboard designs for good emoji entry [36]. An encoder-decoder model named as Seq2Emoji was used to predict multiple emojis based on a short text, considering the correlations between emojis and predicting emoji after the fully connected layer of hierarchical neural networks [37]. The features of given texts and emojis could be automatically extracted by a multi-modal Siamese-based framework [38]. To represent the conceptual content through emoji from cognitive strategies, participants were asked to use emoji to provide semantic representations for abstract and concrete concepts [39].

Some scholars have also studied the emotional information of emoji [40,41]. To learn the final sentiment classifier, Tweets and GitHub posts containing emojis were leveraged to learn text representations through emoji prediction [42]. An emoji sentiment lexicon using an unsupervised sentiment analysis system was constructed to analyze the sentiments expressed by emojis in online textual messages without human annotation [43]. A large-scale data set of emoji usage containing both text-emoji and image-emoji relationships within Tweets was used to predict emoji from both text and images [44]. Taking personal factors and contextual factors into account, a given textual microblog post of a user can be used to predict emojis [45].

Approach

In the task of emotion recognition, the output of neural networks is the corresponding label of the maximum value of the fully connected layer. When the maximum and the second maximum of the output of the fully connected layer are close, only the label corresponding to the maximum is used as the output result and other options are ignored, which will introduce errors. To reduce errors, the corresponding label of the maximum of the fused results of the neural network and the lexicon-based emotional fuzzy inference model are defined as the emotional tendency of the input text.

Sentiment distribution of a document based on neural network

Define the normalized output of the fully connected layer of neural networks as the sentiment membership of the senti-

ment classification task,

$$P_{dp} = [P_{dp_0} \ P_{dp_1} \ P_{dp_2}], \quad (1)$$

where P_{dp_0} is the sentiment membership degree of neural emotion, P_{dp_1} is the sentiment membership degree of positive emotion, P_{dp_2} is the sentiment membership degree of negative emotion and $P_{dp_0} + P_{dp_1} + P_{dp_2} = 1$.

The smallest subtree-based sentiment enhancement reasoning model

To express the emotion of a text more completely, the modifiers in front of the emotion words are regarded as the enhancement of emotion, such as negative words and degree adverbs that are directly related to emotional words. The provided model only discusses 3 consecutive words that modify emotional words and are in the same smallest subtree with the emotional word.

Definition of the smallest subtree of w_i and w_j : in the subtree set of a sentence, there is no subtree that contains both w_i and w_j and has fewer branches and leaves than the existing subtree.

Representation of emotional words

Two tuples (w_i, Po_i) are defined to represent the i th emotional word w_i in a text S . Po_i is emotional polarity of w_i with the definition as follows,

$$\text{Po}_i = \begin{cases} +1, & \text{if positive_emotion} \\ 0, & \text{if neural} \\ -1, & \text{if negative_emotion.} \end{cases} \quad (2)$$

The domain of emotional polarity is defined as

$$P = \{u_0, u_+, u_-\}, \quad (3)$$

where u_0 denotes neural emotion, u_+ denotes positive emotion and u_- denotes negative emotion.

The emotional polarity of the emotional word w_i on the domain P is defined as

$$P_{w_i} = \begin{cases} [0 \ 0 \ 0], & \text{if } \text{Po}_i = 0 \\ [0 \ \text{Po}_i \ 0], & \text{if } \text{Po}_i > 0 \\ [0 \ 0 \ |\text{Po}_i|], & \text{if } \text{Po}_i < 0. \end{cases} \quad (4)$$

Influence of negative words

If continuous negative words and the emotional word w_i are in the same smallest subtree, the emotional polarity takes the opposite number.

$$(w_i, \text{Po}_i) = (w_i, (-1)^{n_{\text{neg}}} \text{Po}_i), \quad (5)$$

Table 2 Examples of some degree adverbs

No.	Chinese	English
D_1	有 点 喜 欢	like slightly
	喜 欢	like
	非 常 喜 欢	like very much
D_{str}	非 常	very much
	特 别	
	尤 其	
	很 超 级	
D_{sli}	一 点 点	slightly
	一 般	
	有 点	
	稍 微	
	弱 弱 的	

where n_{neg} is the number of continuous negative words with direct modification of w_i .

If a negative word, an emotional word, and a question mark are in the same smallest subtree, the emotional polarity Po_i is unchanged,

$$(w_i, \text{Po}_i) = (w_i, (-1)^{1+1} \text{Po}_i) = (w_i, \text{Po}_i). \quad (6)$$

Influence of degree adverbs

In Chinese, adverbs of degree with fuzzy semantics are often used to express different degrees of emotion, such as D_1 listed in Table 2, which are named as degree adverbs. Degree adverbs are stored in two dictionaries D_{str} and D_{sli} according to their semantics, where D_{str} denotes strong emotional modifiers, and D_{sli} denotes slightly emotional modifiers. Some examples of adverbs of degree are shown in Table 2.

An emotional enhancement factor β_{adb} is defined to describe the emotional enhancement of degree adverb v_{word_i} to its adjacent emotional words w_i .

$$\beta_{\text{adb}} = \begin{cases} \beta_{\text{str}}, & \text{if } v_{\text{word}_i} \in D_{\text{str}} \\ \beta_{\text{sli}}, & \text{if } v_{\text{word}_i} \in D_{\text{sli}}, \end{cases} \quad (7)$$

where $\beta_{\text{str}} > 1$ and $\beta_{\text{sli}} < 1$.

If the emotional word is modified by more than one degree adverb, the fuzzy product of all emotional memberships of the degree adverbs is set as the emotion enhancement factor.

$$\beta_{\text{adb}} = \beta_{\text{adb}}^1 \vee \beta_{\text{adb}}^2 \vee \dots \vee \beta_{\text{adb}}^p = \max(\beta_{\text{adb}}^1, \beta_{\text{adb}}^2, \dots, \beta_{\text{adb}}^p), \quad (8)$$

where β_{adb}^p is the p th consecutive emotional enhancement modifier of w_i .

Therefore, the emotional value of w_i is changed as follows,

$$Po_i = \beta_{adb} Po_i. \quad (9)$$

Influence of degree adverbs and negative words

If the modifier in front of the emotional word w_i contains both degree adverbs and negative words, combine Eqs. 4 and 8 to obtain the emotional enhancement representation of w_i ,

$$Po_i = \beta_{adb} (-1)^{n_{neg}} Po_i. \quad (10)$$

Emotional polarity of a sentence

Substituting the Po_i obtained from Eq. 9 into Eq. 3, the sentiment matrix P_{w_i} of (w_i, Po_i) is obtained. The sum of the sentiment value of all sentiment words is defined as the sentiment value of the sentence S ,

$$Po_S = [Po_{S0} \quad Po_{S1} \quad Po_{S2}] = \sum_{i=1}^{n_w} P_{w_i} \quad (11)$$

where n_w is the number of emotional words in S .

Influence of punctuation

Some punctuation, such as question mark, exclamation mark can enhance the emotional expression.

$$Po_{Si} = Po_{Si} (1 - \lambda_{qm} n_{qm}), \quad i = 0, 1, 2, \quad (12)$$

where $\lambda_{qm} (> 0)$ is the emotional decay factor of question marks and n_{qm} is the number of question marks in the end of S . The question mark only strengthens emotional expression and does not change the emotional polarity of a sentence, therefore, $1 - \lambda_{qm} n_{qm} > 0$.

$$Po_{Si} = Po_{Si} (1 + \lambda_{em} n_{em}), \quad i = 0, 1, 2, \quad (13)$$

where $\lambda_{em} (> 0)$ is the emotional enhancement factor of exclamation marks and n_{em} is the number of exclamation marks in the end of S .

Normalize the emotional expression

$$Po_S = [Po_{S0} \quad Po_{S1} \quad Po_{S2}], \quad (14)$$

where $Po_{Si} = \frac{Po_{Si}}{Po_{S0} + Po_{S1} + Po_{S2}}, \quad i = 0, 1, 2$.

Fuzzy recognition-based emotion analysis of emoji

To let emoji express emotions more accurately and comprehensively, membership is introduced to express the various emotional orientation values, which is defined as the emotional membership of emoji. The steps are as follows:

- Utilize the emotional word distribution characteristics of the contextual text of emoji to train the emotional representation of emoji, and normalize the emotional vector. We divide the emotional tendencies of texts into five categories: happiness, like, disgust, sadness, and calm.

$$e_i = [A_1 \quad A_2 \quad A_3 \quad A_4 \quad A_5]^T, \quad (15)$$

where $\sum_{i=1}^5 A_i = 1$, denoting the respective probability of happiness, like, disgust, sadness, and calm.

- Define the value represented by the emoji embedding as the emotional membership degree of the emoji in these five emotions.

$$A(e_i) = \{A_1(e_i), A_2(e_i), A_3(e_i), A_4(e_i), A_5(e_i)\}, \quad (16)$$

where e_i denotes the i th emoji and $A_i(e_i)$ is the i th value in emoji embedding e_i .

If only the maximum emotional tendency of emoticons is considered, the emotional tendency is determined by the maximum value,

$$A(e_i) = \max\{A_1(e_i), A_2(e_i), A_3(e_i), A_4(e_i), A_5(e_i)\}. \quad (17)$$

Equation 17 only considers the maximum emotional membership of emoticons, ignoring the relationship between emoticons and emotional tendencies in different contexts. Equation 16 shows that emojis can express different emotional tendencies in different contexts. The framework of sentiment embedding of emoji is listed in Algorithm 1.

Tree-based fuzzy recognition emotion enhanced reasoning

Emojis belong to different emotional tendencies, which not only reflect the emotional distribution of emojis in the corpus, but also reflect the different emotional expressions of the context. Combining the distribution of emotional words in a text and the fuzzy emotional recognition of emoticons, a tree-based fuzzy recognition emotion enhancement model is established.

For the emotional tendency of emojis, happiness and like are regarded as positive emotions, disgust and sadness are

Algorithm 1 Framework of sentiment embedding of emoji based on emotional word dictionary.

Require: input seed emotional words w_{e_i} of the five emotional word dictionaries D_t , ($t = 1, 2, 3, 4, 5$);

Ensure: compute the sentiment embedding of emoji e_i , P_{e_i} ;

- 1: compute the semantic similarity between the input words and the seed words in each sentiment dictionary to expand the sentiment dictionaries;
- 2: compute the semantic similarity between the emoji e_i and each word w_{e_j} in sentiment dictionary and sum up all the similarities, $S_t \leftarrow \sum_j (Sim(e_i, w_{e_j}))$;
- 3: compute the membership degree of emoji e_i , $A_{e_i} = \{u_t, t = 1, 2, 3, 4, 5\} \leftarrow \{\frac{S_t}{\sum_{t=1}^5 S_t}, t = 1, 2, 3, 4, 5\}$;
- 4: compute the sentiment embedding of emoji e_i , P_{e_i} ;
- 5: **return** P_{e_i} ;

regarded as negative emotions, and calm is regarded as no emotion.

$$A_e(e_i) = \{A_{e0}(e_i), A_{e1}(e_i), A_{e2}(e_i)\}, \quad (18)$$

where $A_{e0}(e_i) = A_5(e_i)$, denoting no emotion; $A_{e1}(e_i) = A_1(e_i) + A_2(e_i)$, denoting a positive emotion; $A_{e2}(e_i) = A_3(e_i) + A_4(e_i)$, denoting a negative emotion.

The emotional tendency of multiple emojis is

$$P_{em} = [A_{e0} \ A_{e1} \ A_{e2}], \quad (19)$$

where $A_{e0} = \sum_{i=1}^{n_e} A_{e0}(e_i)$, $A_{e1} = \sum_{i=1}^{n_e} A_{e1}(e_i)$, $A_{e2} = \sum_{i=1}^{n_e} A_{e2}(e_i)$ and n_e is the number of emojis.

Normalize the emotional tendency of multiple emojis

$$P_{em} = [A_{e0} \ A_{e1} \ A_{e2}],$$

$$A_{ei} = \frac{A_{ei}}{A_{e0} + A_{e1} + A_{e2}}, \quad i = 0, 1, 2, \quad (20)$$

Add Eqs. 14–20 to obtain the emotion of the text containing emojis,

$$P_e = [P_{e0} \ P_{e1} \ P_{e2}] = P_{em} + P_{OS}, \quad (21)$$

where P_{e0} , P_{e1} , P_{e2} respectively denote neural emotion, positive emotion and negative emotion.

Normalize the emotion of the text containing emojis $P_e = [P_{e0} \ P_{e1} \ P_{e2}]$,

$$P_{ei} = \frac{P_{ei}}{P_{e0} + P_{e1} + P_{e2}}, \quad i = 0, 1, 2. \quad (22)$$

Fusion of neural network and lexicon

Combine the deep learning results with lexicon-based results to get the emotional tendency of the text,

$$T_e = \max\{\lambda P_{ei} + (1 - \lambda) P_{dp_i}, i = 0, 1, 2\}, \quad (23)$$

where P_{dp_i} is the output of the full connection layer of deep neural network and λ is the fusion factor, $0 < \lambda < 1$.

Experiment

Datasets and evaluation metrics

The emotional dictionary and emoji embedding in the provided model are the extended dictionary given in [10]. To evaluate the proposed model, macro-precision, macro-recall, macro-F1_score¹ are chosen to evaluate emotion recognition results from Weibo and Douyin.

The Douyin data set contains 26,000 pieces of data with an average sentence length of 13.67 Chinese characters. Three college students manually annotated 3200 texts as a training set. NLPCC2013^{2,3,4} are the public corpus of the Chinese Weibo sentiment analysis competition with an average length of sentences of 38.33 Chinese characters, in which each document is combined with a Weibo topic and its comments. In NLPCC2012,⁵ we treat one Weibo topic as a document. Therefore, the NLPCC2012 dataset consists of some documents with varying length.

Experimental setting and process

The input word2vec and emoji2vec are 100-dimension embeddings, trained on a corpus which consists of 32,000 comments captured in the Douyin APP and 36,000 comments crawled from Sina Weibo. To evaluate the performance of our methods, we combine the calculation results of deep neural networks (DP) with lexicon-based emotion reasoning of texts and emotion distribution of emojis to form different comparison models, as illustrated in Algorithm 2. .

Results

The provided algorithms are listed as follows, where DP is replaced by convolutional neural networks (CNN), long short-term memory (LSTM), bi-directional long short-term memory (BiLSTM) or gate recurrent unit (GRU) in the following experimental results.

1. **DP + emoji** denotes that the emoji embedding is concatenated in the input embedding of DP, and the out-

¹ <http://tcci.ccf.org.cn/conference/2013/dldoc/ev02.pdf>.

² <http://tcci.ccf.org.cn/conference/2013/dldoc/evans02.zip>.

³ <http://tcci.ccf.org.cn/conference/2013/dldoc/evsam02.zip>.

⁴ <http://tcci.ccf.org.cn/conference/2013/dldoc/evdata02.zip>.

⁵ http://tcci.ccf.org.cn/conference/2012/pages/page10_dl.html.

Algorithm 2 Framework of fusion of the output of neural networks and lexicon-based fuzzy inference emotional enhancement model.

Require: input sentence, S ; emotional word dictionary; emoji dictionary; degree adverb dictionary; negative word dictionary

Ensure: Analyze the sentiment polarity of the sentence S , T_e ;

- 1: divide the input document into a text P_t , emojis P_e and punctuation at the end of the sentence;
- 2: input the word2vec and emoji2vec of the document into neural networks in order;
- 3: store the output of the fully connected layer of neural networks for sentiment analysis, P_{dp} ;
- 4: extract the emotional word and its emotional polarity, (w_i, P_{oi}) ;
- 5: generate the subtree set of P_t , T_S ;
- 6: compute the smallest subtree-based emotional enhancement of degree adverbs to the modified emotional words;
- 7: compute the smallest subtree-based emotional enhancement of negative words to the modified emotional words;
- 8: sum up the sentiment distributions of all sentiment words in the sentence to obtain the sentiment distribution of text P_t , P_{oS} ;
- 9: compute emotional enhancement of question mark and exclamation mark to the sentence sentiment;
- 10: compute the sentiment embedding of emojis, P_e ;
- 11: fuse the neural network with lexicon-based fuzzy inference emotion enhancement model, T_e ;
- 12: **return** T_e ;

put result incorporates the fuzzy emotional reasoning enhancement of emojis.

2. **DP + subtree** denotes that the negative words and degree adverbs are considered in the input embedding of DP, and the output result combines the negative words and degree adverbs to enhance the fuzzy emotional reasoning.
3. **DP + all** denotes that the input embedding of DP considers emojis, negative words, degree adverbs, question marks, and exclamation marks, and the output result combines emoji, negative words, degree adverbs, question marks, and exclamation marks to enhance fuzzy emotional reasoning.

Overall results

For the sake of comparison, all parameters of the same neural network for all corpora are set the same in this article and the rest of the parameters except for the neural network on the same corpus are set to the same: $\lambda = 0.7$, $\beta_{str} = 1.2$, $\beta_{sli} = 0.8$ and $\lambda_{qm} = \lambda_{em} = 0.05$. By comparison, it is found that the calculation results of the four neural networks of the provided model are significantly improved in NLPCC2012, NLPCC2013 and Douyin, as shown in Table 3. Especially in the corpus NLPCC2013, the precision of CNN_tree is 22% higher than that of CNN_fr and 31.6% higher than that of CNN_flip. The precision of BiLSTM_tree is 31% higher than that of BiLSTM_fr, reaching the maximum rate of improvement. For the four kinds of neural networks, the precision of the provided model is improved by an average of

27.5% over DP_fr and 26% over DP_flip. The main reason is that the texts of NLPCC2013 are mainly long texts with rich emotional expression and few emojis. However, for the other corpus Douyin with short texts and emoji used frequently to express emotions directly, the improvement rate of the provided model is not as high as that on NLPCC2013. Especially for CNN, LSTM, and BiLSTM, the effects of DP_fr are better than DP_flip. The main reason is that the comments in Douyin are mainly based on the storyline in the video, so the emotional expression is very direct and concise. Commonly, the superpositions of emotional words, emojis, question marks, and exclamation marks are used to highlight strong emotions. Directly counting the number of emotional words and emojis can highlight the expression of emotions. Moreover, an opposite conclusion is reached on NLPCC2012, mainly because there are fewer emojis in sentences in NLPCC2012 compared to Douyin. Therefore, through experimental comparison and reason analysis, the provided model effectively enhances the emotional expression of the text from multiple angles.

Detailed results of different datasets

Influence of different deep learning models on NLPCC2013

Weibo comments are context-based, with free and diverse expressions, including irony, metaphors, slang and other indirect emotional expressions, which makes it difficult for computers to evaluate the sentiment of texts. By comparing the fuzzy emotional reasoning results of different deep learning models, the effects of emoji, degree adverbs and negative words on text emotions are analyzed (Table 4). Compare the effects of corresponding elements by concatenating corresponding element embedding to the input vector of the deep learning model and adding corresponding elements to the input of the inference model. Due to the long length of texts, netizens are accustomed to using emojis at the beginning or at the end of the text to enhance their emotions, which makes the emotional expression of the text information complex and diverse while the emojis are less and single. Therefore, the emotional contributions of emojis in the three deep learning models are lower than the emotion enhancement effects of modified adverbs, as shown in Fig. 2. Comparing the three figures, it is found that the emotional enhancement effect of the modified adverb in the calculated results of LSTM and BiLSTM considering the word order is stronger than the enhanced reasoning of the modified adverb in the CNN without considering the word order. As shown in Fig. 2b, c, the emotional enhancement effects of degree adverbs and negative words are greater than the emotional enhancement effects of emoji, while in Fig. 2a, the opposite is true.

Table 3 Compared results of different models, where DP_fr denotes the model considering the influence of the number of emojis and negative words on document sentiment [10], DP_flip denotes that the sentiment polarity of emotional word is flipped when its adjoint word is a negative word [9] and DP_tree is the provided model

Model	NLPCC2012		NLPCC2013		Douyin	
	Accuracy	Precision	Accuracy	Precision	Accuracy	Precision
CNN_fr	0.808	0.788	0.610	0.652	0.797	0.704
CNN_flip	0.813	0.628	0.587	0.605	0.673	0.650
CNN_tree	0.846	0.857	0.725	0.796	0.828	0.777
LSTM_fr	0.817	0.799	0.646	0.660	0.803	0.714
LSTM_flip	0.829	0.688	0.625	0.637	0.793	0.708
LSTM_tree	0.834	0.834	0.806	0.838	0.841	0.793
BiLSTM_fr	0.825	0.806	0.572	0.550	0.829	0.781
BiLSTM_flip	0.826	0.807	0.609	0.606	0.760	0.775
BiLSTM_tree	0.847	0.847	0.782	0.794	0.839	0.791
GRU_fr	0.791	0.783	0.634	0.615	0.807	0.726
GRU_flip	0.808	0.793	0.659	0.653	0.812	0.777
GRU_tree	0.828	0.830	0.673	0.717	0.843	0.799

Bold indicates the maximum value obtained on each corpus

Table 4 Compared results of different models on NLPCC2013

Metrics	CNN			LSTM			BiLSTM		
	+ emoji	+ subtree	+ all	+ emoji	+ subtree	+ all	+ emoji	+ subtree	+ all
Accuracy	0.6211	0.6742	0.7246	0.5840	0.7888	0.8055	0.5715	0.7167	0.7821
Precision	0.6477	0.7448	0.7960	0.5764	0.7828	0.8375	0.5497	0.7327	0.7941
Recall	0.5363	0.5868	0.6516	0.5273	0.7070	0.7587	0.5276	0.6577	0.7445
F1-score	0.5553	0.6162	0.6867	0.5508	0.7430	0.7860	0.5384	0.6792	0.7634

Comparing the effects of CNN on different datasets

Comparing the performance of inference-based CNN on three different corpora, the experimental effects of the provided model on corpora with different characteristics on emotional expression are analyzed. Comparing the distributions of curves in the three figures illustrated in Fig. 3, we can find that the main ways of expressing emotions in different corpora are different. In Douyin, to express emotions visually and intuitively, the text is short, the emotions are highly targeted, and the expressions of strong emotions are diverse. Exclamation marks, question marks, and emoticons are frequently used. Hence, the influence of emojis on sentiment is higher than the emotion advancement of adverbs, as listed in Fig. 3a. For the same sentiment prediction for short texts, the sentiment impact factors of NLPCC2012 are different from Douyin. In NLPCC2012, the emotional expression of the text is related to the context. There are no emotional words in some texts, but only the emotional expression of the previous comment, such as: same as above, plus one. The emotion of this type of texts is defined as related to the emotions of the previous comments. Moreover, some emoticons in Weibo are transformed from texts, such as: Happy May 1st International Labor Day, [May] [1st]. The combination of [May] [1st] is non-emotional emoticons, but according to

the context it should be positive emotion, and the model cannot recognize its positive emotion. Therefore, the influence of emoji in Weibo is lower than that of subtree, as shown in Figs. 3b, c. Long text has more grammatical modifications than short text, although long text is also composed of short texts. Hence, the contribution of adverbs is higher than emojis, as reflected in Fig. 3b, c (Table 5).

Emotional orientation analysis of Weibo names on NLPCC2012

The sentiment prediction results of the corpus obtained by deleting the Weibo names in the corpus one by one. As the number of Weibo names ignored increases, Precision fluctuates greatly, and Accuracy changes very little, as listed in Table 6. Take $\Delta P_{dp} = 0.3$ for example. Comparing the characteristics of Weibo topic names, it is found that when the subject name describes the object and event with strong pertinence, the emotional tendency of the comments is direct and strong. Thus, the reasoning effect is good, and the corresponding *Precision* is relatively large. When the event described by the topic name cannot be simply judged as correct or wrong in binary, the contents of the comments on Weibo are discrete and there are inconsistencies in the topic

Fig. 2 Detailed influence of different deep learning models on NLPCC2013

of the comments, which makes *Precision* low and forms a trough, as shown in bold in Table 6.

Changing ΔP_{dp} means that the number of sentences enhanced by fuzzy inference has changed. If ΔP_{dp} is reduced, the range of fuzzy inference is reduced, that is, the difference between the output results of the fully connected layer of the deep neural network is reduced, and the ambiguity is increased. When increasing the number of Weibo names ignored, *Precision* of $\Delta P_{dp} = 0.5$ fluctuates first

and then almost stabilizes, while *Precision* of $\Delta P_{dp} = 0.3$ always fluctuates greatly, as illustrated in Fig. 4a. With the increase of the number of Weibo names ignored, the change trends of *Precision* in the two cases are almost the same, as shown in Fig. 4b. The research on Weibo name orientation shows that the neural network does have ambiguity among the output results of the fully connected layer when processing sentiment classification, and this ambiguity does affect the effectiveness of the neural network on emotion predic-

Fig. 3 Detailed influence of CNN on different datasets **Table 5** Compared results of CNN on different datasets

Metrics	NLPCC2013			NLPCC2012			Douyin		
	+ emoji	+ subtree	+ all	+ emoji	+ subtree	+ all	+ emoji	+ subtree	+ all
Accuracy	0.6211	0.6742	0.7246	0.8195	0.8462	0.8462	0.7968	0.6734	0.8278
Precision	0.6477	0.7448	0.7960	0.4525	0.8570	0.8570	0.7040	0.6504	0.7772
Recall	0.5363	0.5868	0.6516	0.4676	0.5632	0.5632	0.7121	0.6076	0.7536
F1-score	0.5553	0.6162	0.6867	0.4599	0.6797	0.6797	0.7080	0.6283	0.7652

Table 6 Emotional orientation of Weibo names on NLPCC2012, where ΔP_{dp} is the difference between the maximum and second maximum of the output of full connected layer of neural networks

ΔP_{dp}	Metrics	Number of Weibo names ignored									
0.3	Precision	1	2	3	4	5	6	7	8	9	10
		0.8479	0.8915	0.7454	0.7955	0.7895	0.8876	0.8831	0.7671	0.8908	0.8397
		11	12	13	14	15	16	17	18	19	20
	Accuracy	0.8960	0.7604	0.8393	0.9025	0.8498	0.8455	0.7648	0.7285	0.8004	0.8481
		1	2	3	4	5	6	7	8	9	10
		0.8430	0.8392	0.8416	0.8488	0.8398	0.8375	0.8402	0.8387	0.8344	0.8402
0.5	Precision	11	12	13	14	15	16	17	18	19	20
		0.8371	0.8425	0.8402	0.8371	0.8394	0.8412	0.8397	0.8420	0.8243	0.8405
	Accuracy	1	2	3	4	5	6	7	8	9	10
		0.8038	0.7981	0.7076	0.8080	0.7160	0.7974	0.8025	0.6868	0.7917	0.7996
		11	12	13	14	15	16	17	18	19	20
		0.7958	0.8033	0.7990	0.7953	0.8044	0.8064	0.8052	0.8063	0.8018	0.8019
	Precision	1	2	3	4	5	6	7	8	9	10
		0.8250	0.8200	0.8214	0.8278	0.8202	0.8198	0.8233	0.8210	0.8156	0.8202
	Accuracy	11	12	13	14	15	16	17	18	19	20
		0.8183	0.8241	0.8202	0.8191	0.8264	0.8274	0.8266	0.8262	0.8247	0.8243

Fig. 4 Comparing the detailed influence of fuzzy reasoning parameter on emotional orientation of NLPCC2012, where delta is the difference between the maximum and second maximum of the output of full connected layer of neural networks, $\text{delta} = \Delta P_{dp}$

Table 7 Detailed results of fuzzy reasoning based CNN

No.	Sentence	Label	P_{cnn}	P_{em}	P_{fr}	Output
P_1	准备带女儿过来玩 🙏🙏 🙏 哎 : 😊😊 I am going to bring my daughter over to play, hey!	NEG	[0.572 0.142 0.286]	[0.14 0.13 0.73]	[0.356 0.136 0.508]	NEG
P_2	不应该加音乐，听不到钟声 🙏 Music should not be added, no bells can be heard.	NEG	[0.543 0.103 0.354]	[0.10 0.42 0.48]	[0.322 0.261 0.417]	NEG
P_3	肯定不是张一白指导的，差点美感 🙏🙏 This scene is definitely not directed by Zhang Yibai, because it lacks aesthetics.	NEG	[0.091 0.576 0.333]	[0.005 0.18 0.815]	[0.048 0.378 0.547]	NEG
P_4	让我们以快乐为本 🌺🌺🌺 Let us take happiness as the basis.	POS	[0.491 0.353 0.156]	[0.33 0.60 0.07]	[0.411 0.506 0.113]	POS

Table 8 Detailed results of fuzzy reasoning based LSTM

No.	Sentence	Label	P_{lstm}	P_{em}	P_{fr}	Output
P_1	准备带女儿过来玩 🙏🙏 🙏 哎 : 😊😊 I am going to bring my daughter over to play, hey!	NEG	[0.428 0.002 0.570]	[0.14 0.13 0.73]	[0.284 0.066 0.650]	NEG
P_2	不应该加音乐，听不到钟声 🙏 Music should not be added, no bells can be heard.	NEG	[0.600 0.005 0.395]	[0.10 0.42 0.48]	[0.35 0.212 0.438]	NEG
P_3	肯定不是张一白指导的，差点美感 🙏🙏 This scene is definitely not directed by Zhang Yibai, because it lacks aesthetics.	NEG	[0.009 0.445 0.546]	[0.005 0.18 0.815]	[0.007 0.313 0.680]	NEG

tion. Therefore, fuzzy inference based on neural network is necessary (Table 7).

Analyze sentences in detail

Take some calculated results of some sentences for example through different deep learning models based on fuzzy inference in detail, as illustrated in Tables 5, 8, and 9. Take P_1 calculated by CNN as an example to explain the calculation process of the output result.

1. Obtain the emoji embedding P_{em1} of 🙏🙏🙏, $P_{em1} = [0.26 \ 0.13 \ 0.61]$.
2. Similarly, obtain the emoji embedding P_{em2} of 😊😊 by Algorithm 1, $P_{em2} = [0.02 \ 0.26 \ 0.85]$.

3. Add up all the sentiment embeddings and normalize to obtain the sentiment embedding P_{em} of emojis, $P_{em} = [0.14 \ 0.13 \ 0.73]$.
4. Fuse the emotional membership of the fully connected layer of CNN with the sentiment embedding of emojis to obtain the emotional classification representation of the text, $P = \lambda P_{em} + (1 - \lambda) P_{cnn} = [0.356 \ 0.136 \ 0.508]$ with the condition of $\lambda = 0.5$.
5. The sentiment polarity corresponding to the maximum value 0.508 is the output result, NEG.

Table 9 Detailed results of fuzzy reasoning-based BiLSTM

No.	Sentence	Label	P_{bilstm}	P_{em}	P_{fr}	Output
P_1	准备带女儿过来玩 🙏🙏 🙏 哎 : 🙄🙄 I am going to bring my daughter over to play, hey!	NEG	[0.562 0.027 0.411]	[0.14 0.13 0.73]	[0.351 0.078 0.570]	NEG
P_2	不应该加音乐，听不到钟声 🕒 Music should not be added, no bells can be heard.	NEG	[0.549 0.051 0.400]	[0.10 0.42 0.48]	[0.325 0.235 0.44]	NEG
P_3	肯定不是张一白指导的，差点美感 🙄🙄 This scene is definitely not directed by Zhang Yibai, because it lacks aesthetics.	NEG	[0.017 0.538 0.445]	[0.005 0.18 0.815]	[0.011 0.359 0.63]	NEG

Conclusion

In the emotion recognition task, the deep learning model uses the label corresponding to the maximum value in the output result of the fully connected layer as the output result. In a multi-classification task, when the boundary between the maximum and the second maximum of the output result of the fully connected layer is not clear, only considering the maximum value and ignoring other values will introduce noise terms. To reduce the errors caused by the unclear classification, a fuzzy classification membership function of the fully connected layer of the neural network, an emotional membership of emojis and a dictionary-based fuzzy inference model are combined to establish a multi-modal and multi-scale emotion-enhanced inference model based on fuzzy recognition. The emotional tendency of emoji based on the corpus, the emotional enhancement of negative words and degree adverbs to modified emotional words, and the sentiment enhancement of the text by question marks and exclamation marks are discussed. Experiments show that this model can accurately recognize text emotions, especially in the effect of long text emotion recognition on Weibo, which is significantly improved.

Acknowledgements The authors would like to thank the referee for his valuable remarks. This work was supported by the National Natural Science Foundation of China (Grant no. 12171065 and 11671001) and the Doctor Training Program of Chongqing University of Posts and Telecommunications, China (Grant no. BYJS201915 and BYJS201907).

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflicts of interest regarding the publication of this paper.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Wang X, Kou L, Sugumaran V et al (2020) Emotion correlation mining through deep learning models on natural language text. *IEEE Trans Cybern* 99:1–14
- Pavan MC, Dos Santos VG, Lan AGJ et al (2020) Morality classification in natural language text. *IEEE Trans Affect Comput* 99:1
- Ayari N, Abdelkawy H, Chibani A et al (2020) Hybrid model-based emotion contextual recognition for cognitive assistance services. *IEEE Trans Cybern* 1–10
- Gao Y, Gong M, Xie Y et al (2021) An attention-based unsupervised adversarial model for movie review spam detection. *IEEE Trans Multimed* 23:784–796
- Yao X, She D, Zhang H et al (2021) Adaptive deep metric learning for affective image retrieval and classification. *IEEE Trans Multimed* 23:1640–1653
- Yang X, Feng S, Wang D et al (2020) Image-text multimodal emotion classification via multi-view attentional network. *IEEE Trans Multimed* 99:1
- Firdaus M, Chauhan H, Ekbal A et al (2020) EmoSen: generating sentiment and emotion controlled responses in a multimodal dialogue system. *IEEE Trans Affect Comput* 99:1–1
- Deng J, Ren F (2020) Multi-label emotion detection via emotion-specified feature extraction and emotion correlation learning. *IEEE Trans Affect Comput* 99:1–1
- Wu O, Yang T, Li M et al (2020) Two-level LSTM for sentiment analysis with lexicon embedding and polar flipping. *IEEE Trans Cybern* 99:1–13

10. Yan R, Yu Y, Qiu D (2021) Emotion-enhanced classification based on fuzzy reasoning. *Int J Mach Learn Cybern* 1–12
11. Baaj I, Poli J P (2019) Natural language generation of explanations of fuzzy inference decisions. In: 2019 IEEE international conference on fuzzy systems (FUZZ-IEEE), pp 1–6. <https://doi.org/10.1109/FUZZ-IEEE.2019.8858994>
12. Paul S, Ray K, Saha D (2020) Application of fuzzy answer set programming in hybrid deep neural-symbolic architecture. In: 2020 IEEE international conference for convergence in engineering, pp 105–109. <https://doi.org/10.1109/ICCE50343.2020.9290604>
13. Lu X, Bai Y (2019) A new rule reduction method for fuzzy modeling. *IEEE Trans Fuzzy Syst* 28(11):3023–3031
14. Zhou H, Ying H, Zhang C (2020) Effects of increasing the footprints of uncertainty on analytical structure of the classes of interval type-2 Mamdani and TS fuzzy controllers. *IEEE Trans Fuzzy Syst* 27(9):1881–1890
15. Ali M, Adnan M, Tariq M et al (2021) Load forecasting through estimated parametrized based fuzzy inference system in smart grids. *IEEE Trans Fuzzy Syst* 29(1):156–165
16. Gupta P, Kumar S, Suman RR et al (2020) Sentiment analysis of lockdown in India during COVID-19: a case study on Twitter. *IEEE Trans Comput Soc Syst* 99:1–11
17. Gui L, Jia L, Zhou J et al (2020) Multi-task learning with mutual learning for joint sentiment classification and topic detection. *IEEE Trans Knowl Data Eng* 99:1–1
18. Shobana J, Murali M (2021) An efficient sentiment analysis methodology based on long short-term memory networks. *Complex Intell Syst* 7:2485–2501
19. Shekhar S, Garg H, Agrawal R et al (2021) Hatred and trolling detection transliteration framework using hierarchical LSTM in code-mixed social media text. *Complex Intell Syst* 99:1
20. Mai S, Hu H, Xu J et al (2020) Multi-fusion residual memory network for multimodal human sentiment comprehension. *IEEE Trans Affect Comput* 99:1
21. Huang M, Xie H, Rao Y et al (2020) Lexicon-based sentiment convolutional neural networks for online review analysis. *IEEE Trans Affect Comput* 99:1–1
22. Zhou T, Cao J, Zhu X et al (2020) Visual-textual sentiment analysis enhanced by hierarchical cross-modality interaction. *IEEE Syst J* 99:1–12
23. Krishnan PT, Joseph Raj AN, Rajangam V (2021) Emotion classification from speech signal based on empirical mode decomposition and non-linear features. *Complex Intell Syst* 7(4):1919–1934
24. Ma K, Jiang B (2020) Voice of urban park visitors: exploring destination attributes influencing behavioural intentions through online review mining. *Complex Intell Syst* 99:1–1
25. Xu N, Mao W, Wei P et al (2020) MDA: multimodal data augmentation framework for boosting performance on image-text sentiment/emotion classification tasks. *IEEE Intell Syst* 99:1–1
26. Poria S, Hazarika D, Majumder N et al (2020) Beneath the tip of the iceberg: current challenges and new directions in sentiment analysis research. *IEEE Trans Affect Comput* 99:1–1
27. Wang L, Niu J, Yu S (2020) SentiDiff: combining textual information and sentiment diffusion patterns for Twitter sentiment analysis. *IEEE Trans Knowl Data Eng* 32(10):2026–2039
28. Xu J, Li Z, Huang F et al (2020) Visual sentiment analysis with social relations-guided multiattention networks. *IEEE Trans Cybern* 99:1–13
29. Koh JI, Cherian J, Taele P et al (2019) Developing a hand gesture recognition system for mapping symbolic hand gestures to analogous emojis in computer-mediated communication. *ACM Trans Interact Intell Syst* 9(1):134–168
30. Li M, Chng E, Chong A et al (2019) An empirical analysis of emoji usage on Twitter. *Ind Manag Data Syst* 119(8):1748–1763
31. Tomihira T, Otsuka A, Yamashita A et al (2020) Multilingual emoji prediction using BERT for sentiment analysis. *Int J Web Inf Syst* 16(3):265–280
32. Robertson A, Magdy W, Goldwater S (2020) Emoji skin tone modifiers: analyzing variation in usage on social media. *ACM Trans Soc Comput* 3(2):11:1–11:25
33. Wijeratne S, Saggion H, Kiciman E et al (2020) Emoji understanding and applications in social media: lay of the land and special issue introduction. *ACM Trans Soc Comput* 3(2):6:1–6:5
34. Wiegand M, Ruppenhofer J (2021) Exploiting emojis for abusive language detection. In: Proceedings of the 16th conference of the European chapter of the association for computational linguistics, pp 369–380. <https://doi.org/10.18653/v1/2021.eacl-main.28>
35. Kim J, Gong T, Kim B et al (2020) No more one liners: bringing context into emoji recommendations. *ACM Trans Soc Comput* 3(2):9:1–9:25
36. Pohl H, Domin C, Rohs M (2017) Beyond just text: semantic emoji similarity modeling to support expressive communication. *ACM Trans Comput Hum Interact* 24(1):1–42
37. Peng D, Zhao H (2021) Seq2Emoji: a hybrid sequence generation model for short text emoji prediction. *Knowl Based Syst* 214(2):106727
38. Suman C, Saha S, Bhattacharyya P et al (2021) Emoji helps! A multi-modal Siamese architecture for Tweet user verification. *Cogn Comput* 13(2):261–276
39. Wicke P, Bolognesi M (2020) Emoji-based semantic representations for abstract and concrete concepts. *Cogn Process* 21(4):615–635
40. Lou Y, Zhang Y, Li F et al (2020) Emoji-based sentiment analysis using attention networks. *ACM Trans Asian Low Resour Lang Inf Process* 19(5):64:1–64:13
41. Boy S, Ruiter D, Klakow D (2021) Emoji-based transfer learning for sentiment tasks. In: Proceedings of the 16th conference of the European chapter of the association for computational linguistics, pp 103–110. <https://doi.org/10.18653/v1/2021.eacl-srw.15>
42. Chen Z, Cao Y, Yao H et al (2021) Emoji-powered sentiment and emotion detection from software developers' communication data. *ACM Trans Softw Eng Methodol (TOSEM)* 30(2):18:1–18:48
43. Fernandez-Gavilanes M, Juncal-Martinez J, Garcia-Mendez S et al (2018) Creating emoji lexica from unsupervised sentiment analysis of their descriptions. *Expert Syst Appl* 103(aug.):74–91
44. Cappallo S, Svetlichnaya S, Garrigues P et al (2019) New modality: emoji challenges in prediction, anticipation, and retrieval. *IEEE Trans Multimed* 21(2):402–415
45. Zhao G, Liu Z, Chao Y et al (2020) CAPER: context-aware personalized emoji recommendation. *IEEE Trans Knowl Data Eng* 99:1–1

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.