IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 25 May 2024, accepted 10 June 2024, date of publication 18 June 2024, date of current version 26 June 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3416379

RESEARCH ARTICLE

TaneNet: Two-Level Attention Network Based on Emojis for Sentiment Analysis

QIN ZHAO[1,2], (Senior Member, IEEE), PEIHAN WU[1], JIE LIAN[1], DONGDONG AN[1], AND MAOZHEN LI[3]

[1]Shanghai Engineering Research Center of Intelligent Education and Big Data, Shanghai Normal University, Shanghai 200234, China
[2]Key Laboratory of the Ministry of Education for Embedded System and Service Computing, Tongji University, Shanghai 200090, China
[3]Department of Electronic and Computer Engineering, Brunel University London, UB8 3PH London, U.K.

Corresponding author: Dongdong An (andongdong@shnu.edu.cn)

This work was supported in part by the National Key Research and Development Program of China under Grant 2022YFB4501704; in part by the National Natural Science Foundation of China under Grant 62302308, Grant U2142206, Grant 62372300, and Grant 61702333; and in part by Shanghai Sailing Program under Grant 21YF1432900.

ABSTRACT During online communication, users often use irregular and ambiguous words, and sometimes use irony to express sarcasm. These words are difficult to analyze through text analysis, which poses a significant challenge for text sentiment analysis. As a novel communication method, emojis have a significant correlation with user emotions. In this paper, we use emojis to analyze the sentiment of short texts. Firstly, we validate that user information can help reduce the uncertainty of some emojis and use this information to identify the polarity of emojis. Then, we generate emoji representations by merging positional information, semantic information, emotional information, and frequency of appearance. Furthermore, we propose TaneNet, a two-level attention network based on emojis, which combines clause vectors and emoji representations to study the impact of emojis on the emotions of each clause in the text. Empirical results on two real-world datasets demonstrate that TaneNet outperforms existing state-of-the-art methods.

INDEX TERMS Emojis, attention mechanisms, word embedding, sentiment analysis, neural network.

I. INTRODUCTION

Social networks have eliminated spatial limitations and enabled people to communicate and provide real-time feedback regardless of their distance from each other, making daily communication more convenient and promoting timely exchange of information. With the changing communication trends, social media users have shifted to using short-text platforms like microblogs and Twitter to express their thoughts in a quick and concise manner, leading to increased participation in information sharing. Short-text communication, limited to 140 words, allows individuals to express their emotions and opinions in a manner similar to actual conversations [1]. The short-text is usually relatively short, and the grammatical structure is not standardized, but it contains rich emotional tendencies of users. Consequently, short-text sentiment analysis has become increasingly important for understanding social network dynamics. However, the use of casual language including slangs, acronyms, and internet jargon [2] poses significant challenges to short-text sentiment analysis, spurring research in developing more effective methods from diverse angles.

Emojis [3], visual shorthand symbols, have been in use on social media for many years and saw a significant increase in usage and standardization following the Unicode Consortium's release of Unicode 6.0 in 2011. This standardization has made emojis a popular and accurate way for individuals to express their sentiments across various digital platforms. The inclusion of emoji in plain text has proven to be an effective means of enhancing sentiment expression and avoiding misunderstandings. In modern social media communication, emojis not only enrich the dimensionality of emotional expression but also offer a solution to the challenges posed by sarcasm and ambiguity in textual analysis. Studies have shown that the accuracy of sentiment analysis significantly improves when text is combined with emojis, highlighting

The associate editor coordinating the review of this manuscript and approving it for publication was Mostafa M. Fouda.

the importance of integrating visual symbols into sentiment analysis tools. Moreover, the widespread use of emojis reflects the diversity and complexity of emotional expression across different cultures, ages, and genders, which are often inadequately captured by text-only analysis.

While previous research has demonstrated the potential to enhance sentiment analysis by integrating emojis, these initial models exhibited specific limitations, particularly in handling the diversity and complexity of microblogging content. These models were also limited in their ability to accurately interpret the nuanced sentiments expressed through emojis, especially when emojis are used in varied contexts that may convey different emotional undertones. Recognizing these challenges, researchers have continued to explore how emojis can be more effectively incorporated into sentiment analysis to improve accuracy and comprehensiveness.

Building on these observations, researchers such as Hogenboom et al. [4] have noted that emojis play a key role in short-text sentiment analysis, as they typically depict the true sentiment of the author. Gupta et al. [5] examine the use of emojis in sentiment analysis, comparing the frequency of emoji usage between top male and female Twitter personalities, and conclude that sentiment analysis is more accurate and comprehensive when combining text and emojis. Amrullah et al. [6] conduct experiments and the results showed that when doing sentiment analysis, the one that includes emoji gets slightly better overall performance than other variants. In support of this claim, there are several examples of emoji use in social media contexts.

Example 1: Huh? Why do I only have slightly more than 15,000 fans for this account, but I have the ability to block and ban others? Awesome! Thanks! 🙂

Example 2: I am so distressed 😭 😭 😭, because I threw the key in when I put my old clothes in the recycling bin, so silly! 🙂

Example 3: Too difficult! I spent an entire year writing this paper and I finally finished it today. Give myself a 👍!

Example 4: Thanks to Thiago for letting me know what it's like to do an overnight check.🙂

As demonstrated in **Example 1**, even though the text incorporates negative terms, the overarching sentiment expressed is positive. This example underscores the limitations of relying exclusively on textual content for sentiment analysis. By observing the consistent use of emoji 🙂 to express positivity, it becomes plausible to infer the underlying positive sentiment, demonstrating the critical role emojis play in enhancing textual sentiment analysis, particularly in ambiguous or ironic expressions. Hakami et al. posit that there exist disparities in expression habits and connotations across genders, underscoring the complexity of interpreting emoji usage [7]. The challenge arises in deciphering the polarity of emojis, as their meanings can vary among different users. Particularly among adolescents, there is a growing trend of employing emojis either independently or as substitutes for text (e.g., **Example 3**), and even resorting to repetitive emoji sequences (as illustrated in **Example 2**) to amplify sentiment intensity [8]. Generally, the presence of emojis in a message can provide valuable insights into the user's sentiment, provided that the polarity of the emojis can be discerned.

Similarly, **Example 4** highlights a subtle yet impactful usage of emojis. "Thanks to Thiago for letting me know what it's like to do an overnight check" uses a positive emoji 🙂 in a context that suggests a less than positive experience, reflecting a nuanced expression common among younger users. This ironic use of emojis presents additional challenges for sentiment analysis, as the traditional text-based approaches may interpret these symbols at face value, leading to misinterpretations.

Conventional methods of text analysis are limited in their ability to discern the extent to which individual sentences contribute to the overall emotional sentiment of the text. As a result, researchers have utilized sentence-level attention mechanisms to parse texts into individual sentences based on punctuation, such as periods [9]. However, this approach is not entirely applicable to sentiment analysis of short texts, which often consist of only one sentence, as demonstrated in **Example 2**. In such cases, it can be difficult to isolate the most useful phrases or terms for sentiment analysis. Thus, identifying relevant and informative textual segments for sentiment analysis in such contexts proves challenging.

Additionally, it should be noted that the conventional principle of attention mechanism is predominantly confined to text analysis, neglecting the potential influence of emojis. Existing research has highlighted that emojis frequently occur in clauses that convey strong emotions, such as **Example 1** (Great! Thank! 🙂) and **Example 2** (I am so distressed 😭 😭 😭), exhibiting their efficacy in reflecting genuine sentiments. To address this issue, we have introduced the relevant properties of emojis to depict them, and introduced a novel definition of clauses that is more succinct than that of sentences [10].

Recognizing the limitations of traditional sentiment analysis methods in adequately capturing the emotional nuances expressed through emojis, particularly in diverse and complex digital communications, we developed TaneNet. This novel framework advances beyond basic emoji-based strategies by incorporating a sophisticated two-level attention mechanism, specifically designed to enhance the processing of the interplay between text and emojis. TaneNet's approach focuses on both word-level and clause-level elements, enabling a more effective capture and interpretation of sentiments in short texts, tailored specifically to the nuances of the platform and the nature of the content being analyzed.

The main contributions of this paper are summarized as follows:

- We present a novel approach for a comprehensive representation of emojis by incorporating multiple dimensions. Specifically, we propose leveraging users' information to determine the sentiment polarity of

emojis. Furthermore, we integrate the sentiment vector, frequency vector, position vector, and semantic vector of emojis to enhance their modeling capabilities.
- Introducing TaneNet, our proposed two-level attention network, which leverages emojis to predict the sentiment polarity of each document. TaneNet incorporates word-level and clause-level attention mechanisms, allowing it to capture fine-grained details and contextual information for more accurate sentiment prediction.
- The experimental results demonstrate the effectiveness of the proposed method. Through comparisons with state-of-the-art methods in sentiment analysis tasks, our model exhibits superior performance, surpassing them in terms of accuracy and overall effectiveness.

The rest of the paper is organized as follows. Section II describes some related work about sentiment analysis and advanced methods. In Section III, we analyze the emoji data and propose our model. The experimental results and the influencing factors are discussed in Sections IV. Finally, conclusions are described in Section V.

II. RELATED WORK

A. SENTIMENT ANALYSIS

The primary objective of sentiment analysis is to classify the sentiment polarity of textual data, which is generally approached as a classification task. Existing sentiment analysis techniques can be classified into supervised, semi-supervised, and unsupervised learning methods, depending on the availability of labeled training data.

Unsupervised learning methods include the development of an emoticon lexicon that leverages the correlation between emotions and emoticons, which is integrated with a standard sentiment lexicon for sentiment analysis [4]. Huang et al. manually constructed an emoticon lexicon and applied the latent Dirichlet allocation algorithm to uncover explicit topics and sentiment [11].

Semi-supervised learning methods include Khan et al. propose a hybrid approach that fuses a vocabulary-based approach with machine learning techniques to improve sentiment analysis accuracy [12], modifying sentiment scores from SentiWordNet [13] using mathematical models. Zhang et al. propose a semi-supervised STCS lexicon model based on SC for accurate sentiment identification of topic-related words [14]. Nuha and Lin employ semi-supervised learning to effectively mine emotional information [15], [16].

Supervised learning methods encompass the research on aspect targets at radical, character, and word levels, combining two levels for modeling [17]. Pang et al. utilize various feature representation techniques and machine learning algorithms for combined comparison experiments [18]. Rao et al. propose a novel neural network with two hidden layers to capture comprehensive emotional information from extended time steps [19]. Additionally, Alexandridis et al. explore aspect-based sentiment analysis using a hybrid deep learning architecture that combines bi-directional LSTM, convolutional layers, and attention mechanisms, enhanced by metadata such as emoji frequency and hashtags, significantly improving the detection of negative sentiments [20]. Routray et al. investigate the applicability of different supervised machine learning approaches for sentiment analysis of students' subjective feedback [21].

Emojis are ubiquitous in microblogs and can better convey emotions. Singh et al. propose a multilabel emoji prediction system for tweets [22]. Researchers have developed sentiment analysis methods incorporating emojis, treating them as features and investigating their impact [23], [24], or training emoji with context using BERT for text feature extraction [25]. However, these methods analyze emojis in isolation, failing to account for their impact on text. We contend that emoji information, such as location and frequency, could significantly contribute to text sentiment analysis.

Some works treat emojis as weakly sentiment labels, assigning fixed sentiment polarity [4], [11]. Li et al. separate emoji and text sentiment for category calculation [26]. However, these works neglect emoji ambiguity, as different individuals may interpret emojis differently. To address this, we employ user information to eliminate emoji ambiguity.

B. DEEP LEARNING

In recent years, deep learning has become mainstream. Zhao et al. [27] use deep neural networks to add additional information to alleviate the data sparsity problem. Yin et al. [28] compared CNNs and RNNs on various tasks in Natural Language Processing. Ling et al. [29] propose a new hybrid neural network model structure, which integrates CNN and Bi-LSTM to extract the deep semantic information of the text, so as to deal with the phenomenon of polysemy and the topic confusion of Sina Weibo. Li et al. [30] propose a framework based on the emotional recurrent unit (ERU), a recurrent neural network that contains a generalized neural tensor block (GNTB) and a two-channel feature extractor (TFE) to tackle conversational sentiment analysis. Wang et al. [31] propose a tensor-based long-short-term memory (LSTM) network, which uses neural networks to extract abundant multimodal and intra-model information formations. It can be seen from the experiment results [32], [33] that the method of combining GRU in the deep learning model has good performance.

Our work employs a bi-directional gated recurrent units (GRU) network that considers the contextual relationship, enabling it to capture information both preceding and following the current time step.

C. ATTENTION MECHANISM

The attention mechanism, renowned for its ability to discern salient features, stands as an efficacious tool particularly suited for research domains entwined with language or imagery. Xu et al. [34] pioneered the integration of attention mechanisms into natural language processing, catalyzing extensive scholarly exploration thereafter. Building upon this foundation, Cao et al. [35] devised a two-level attention model, adept at capturing intricate representations by amalgamating two data embeddings at both the data sample and feature levels.

In the realm of document-level sentiment analysis, pioneering works by Yang et al. [36] and Xu et al. [37] introduced hierarchical attention networks capable of discerning emphasis and elucidating the structural underpinnings of textual content. Chandio et al. [38] further advanced sentiment analysis methodologies by proposing a deep recurrent architecture–RU-BiLSTM–augmented with word embedding and attention mechanisms.

Moreover, Tan et al. [39] and Suman et al. [40] expanded the purview of attention mechanisms by integrating emoticon information into the model, thereby achieving notable performance enhancements. Nevertheless, it is imperative to note that traditional sentence-level attention mechanisms encounter limitations in the context of short text analysis.

Our work divides sentences into shorter length sentences called clauses, and use two-layers attention mechanism which is word level and clause level, adding the emojis representation to clause level to make the sentiment analysis more accurate.

III. METHODOLOGY

In both online and offline contexts, individuals often use ironic statements to express their emotions, utilizing positive emotion words to convey negative sentiments. For instance, as in **Example 4**, the isolated word "Thank" may convey a positive emotion, but in the context of this message, it conveys a negative sentiment.

Not only do sentences have varying polarities and uncertain words, but irregular phrasing can cause important sentiments to be missed. To tackle these challenges, our methodology begins with a critical step of data processing. In this phase, we conduct extensive preprocessing of the data where we use microblogs to train a Word2Vec model to generate word embeddings and manually label some microblogs as our dataset. This foundational step ensures that the subsequent analysis is built on a robust dataset optimized for our specific research needs.

Following data processing, we propose using emojis as a means of analyzing web sentences. Consider **Example 3**, where traditional text analysis might attribute a wrong negative sentiment to the last sentence of the text due to the lack of a key subject, thereby improperly emphasizing "Too difficult". However, by integrating emojis into text analysis, we can obtain more accurate sentiment analysis results.

Therefore, we intend to analyze and represent useful emoji features for sentiment. Subsequently, we fuse the resulting emoji representations with the text embeddings of two layers. Our methodology is shown in Fig.1.

In this paper, the sentiment lexicon used is built by Huang et al. [10]. This lexicon is created by integrating and adjusting three established Chinese sentiment lexicons. The integration and adjustment process ensures that the sentiment labels assigned to emojis are accurate and consistent with the Chinese language context.

TABLE 1. Statistical information of lexicons.

Type	Hownet	Tsinghua	NTUSD	Basic	Seed
#Pos	4566	5567	2810	11184	100
#Neg	4370	4468	8278	14925	100
#Total	8930	10035	11088	26109	200

A. DATA PROCESSING

1) CORPUS CONSTRUCTION

Our corpus was established by retrieving approximately 1.1 million microblogs via the Sina Weibo API and amassing an additional 5 million microblogs from various online sources. The processing steps for these microblogs involved several stages: initially, filtering out posts containing keywords related to advertising and spam; secondly, excising links, hashtags, '@' symbols, and kaomojis, as they are irrelevant to our methodologies. From the filtered microblogs, we selected 11,482 entries containing user information. These were manually annotated as positive or negative, forming the dataset for this study.

2) WORD EMBEDDING

To extract more abstract textual information, we opted for the Word2Vec model [41], specifically employing the Skip-gram technique over the continuous bag-of-words (CBOW) model, following recommendations in [42]. Each microblog was tokenized using the jieba tool, after which stopwords, including punctuation and other meaningless words, were removed, leaving a corpus of 110 million words. We utilized the Word2Vec implementation from the Gensim library, setting the window size to 5, iteration count to 35, and dimensions to 300. Additionally, we set a minimum count threshold of 5 to filter out infrequent words. To expedite the training phase, hierarchical softmax was used instead of negative sampling.

3) SENTIMENT LEXICON CONSTRUCTION

A novel sentiment lexicon was developed by integrating and modifying three existing Chinese sentiment lexicons: Hownet, NTUSD, and Tsinghua. Initially, we purged outdated terms from these lexicons and manually incorporated several contemporary cyber terms. Subsequently, we identified and selected 100 positive and 100 negative words to create a sentiment seed lexicon, characterized by words with more pronounced sentiment relevance. The statistical details of these lexicons are presented in Table 1.

B. FEATURE EXTRACTION

In the presented **Example 1**, the co-occurrence of two polar sentiment words, "block and ban" and "Awesome! Thanks!", poses a challenge in determining the overall

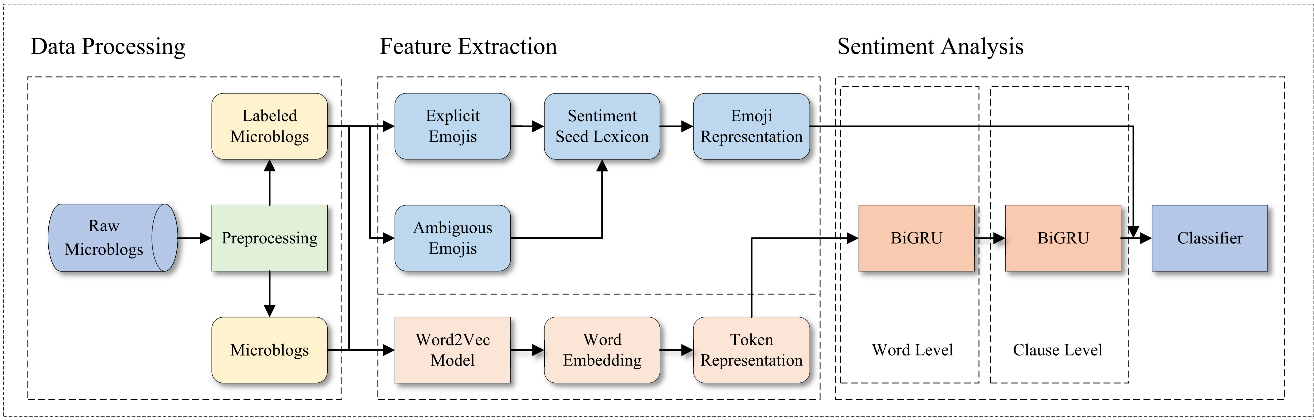


FIGURE 1. Overview of the proposed method.

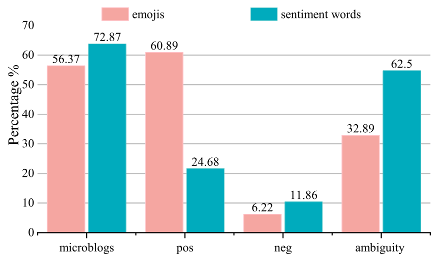


FIGURE 2. Emoji and sentiment word counts.

sentiment of the sentence. However, the presence of the 🙂 emoji after "Awesome! Thanks!" can provide valuable insight into the sentiment of the sentence. By considering the polarity of the emoji, one can gain a better understanding of the overall sentiment, since presuming that the emoji is closer to the positive word would highlight the importance of "Awesome! Thanks!". The explanation is also valid for **Example 2**, albeit more complex, given the presence of two types of emoji in different positions. Notably, the repetition of 😭 determines the overall sentiment of the sentence to be negative. These examples underscore the importance of emojis in sentiment analysis, highlighting the need to consider user habits for emoji feature mining.

It is worth noting that the same emoji 🙂 is used in both **Example 1**, **Example 2**, and **Example 3** with different meanings. This raises the question of whether emojis have the same ambiguity as words, which makes determining the emotional polarity of emojis equally difficult. To address this question, we conducted an experimental analysis in which we compared emojis with sentiment words. The results, presented in Fig. 2, lead to the following conclusions: 1) more than half of the microblogs contain emojis; 2) most emojis are positive; 3) emojis have less ambiguity than sentiment words, and 4) there are still some ambiguous emojis in terms of sentiment. Therefore, to handle these ambiguous cases, we will discuss the treatment of explicit and ambiguous emojis separately.

TABLE 2. Typical explicit emojis.

Sentiment	#Emojis	Typical emojis
Positive	20	🥰 😋 😊 😆 😜 😚 🤗
Negative	20	🥺 😓 😑 😡 🤒 😢 😵

1) EXPLICIT EMOJIS

To differentiate between ambiguous and explicit emojis, we assign sentiment polarities to each category separately. For explicit emojis, we employ a cosine similarity approach to select them based on the comparison between words associated with emojis and words from a sentiment seed lexicon. We utilize pre-trained word embeddings to represent the words. Specifically, we calculate the sentiment score, denoted as $score(e_i)$, for each explicit emoji e_i using the following formula:

$$\mathrm{score}(e_i) = \left| \frac{\sum_{j=1}^{m} \cos(e_i, \mathrm{pos}_j)}{m} \right| - \left| \frac{\sum_{k=1}^{n} \cos(e_i, \mathrm{neg}_k)}{n} \right|, \quad (1)$$

where m, n denotes the number of positive word and negative word in sentiment seed lexicon respectively, in this paper $m = n = 100$, pos_j and neg_k denotes j-th positive word and k-th negative word in sentiment seed lexicon respectively. Finally, we choose the top 20 emojis as positive emojis and the last 20 emojis as negative emojis. Specifically, we list some emojis in Table 2.

2) AMBIGUOUS EMOJIS

According to Huang et al. [10], their research demonstrates that the sentiment associated with an emoji remains consistent for a specific user, despite potential variations across different users. This finding suggests that it would be beneficial to determine the sentiment polarity of an ambiguous emoji specifically for an individual user. By considering the user's historical usage and associated sentiment patterns, we can enhance the accuracy of sentiment analysis for emojis. In this paper, ambiguous emojis are assumed to have different sentiment polarities for different users. This means that the sentiment associated with an ambiguous emoji can vary depending on the user's perspective or interpretation. The sentiment polarity classification of ambiguous emojis takes into account the individual user's sentiment orientation and their co-occurrence patterns with explicit positive and negative emojis.

The Sentiment Orientation PMI (SO-PMI) is used in this context to determine the sentiment polarity of ambiguous emojis. The formula for calculating the SO-PMI of emoji e_i with respect to user u is given by:

$$\text{SO-PMI}_u(e_i) = \frac{\sum_p \text{PMI}(e_i, p)}{n_{\text{pos}}} - \frac{\sum_n \text{PMI}(e_i, n)}{n_{\text{neg}}}, \quad (2)$$

where n_{pos}, n_{neg} respectively denote the number of explicit positive emojis co-occur with e_i and the number of explicit negative emojis co-occur with e_i, PMI is the Pointwise Mutual Information between two emojis. We express the set of explicit positive emojis as UP, and the set of explicit negative emojis as UN, $p \in UP$ and $n \in UN$.

Based on the SO-PMI value, the sentiment polarity of emoji e_i with respect to user u can be determined as follows:

$$\text{class}_u(e_i) = \begin{cases} 1, & \text{SO-PMI}_u(e_i) \geq 0, \\ 0, & \text{SO-PMI}_u(e_i) < 0, \end{cases} \quad (3)$$

If the SO-PMI value is greater than or equal to zero, the emoji is classified as having a positive sentiment polarity. Otherwise, if the SO-PMI value is less than zero, the emoji is classified as having a negative sentiment polarity.

3) EMOJI REPRESENTATIONS

According to the research by Hogenboom et al. [4], emojis are more likely to influence the sentence they appear in. Therefore, this paper also takes into account the positional information of emojis. The positional information is based on the distance between the emoji and the first word in the sentence. Additionally, there is a phenomenon where the same emoji is used multiple times to intensify sentiment expression. Hence, the frequency of emoji occurrences is also considered as a feature. Moreover, the text descriptions of emojis are taken into account as their semantic features. For example, the text description "smile" is used as the semantic feature for the emoji 🙂.

To represent these features, an emoji can be represented as a vector $v_i = \{t_i, p_i, c_i, s_i\}$, where t_i represents the sentiment type of the emoji, indicating its positive or negative sentiment, p_i represents the position feature vector, c_i represents the count of the emoji, and s_i represents the sentiment feature vector, which is derived from the word embedding of the text description associated with the emoji.

C. MODEL CONSTRUCTION

Our proposed model, **TaneNet**, is a Two-level Attention Network based on Emojis. It incorporates a two-layer attention mechanism where not only the word-level but also the clause-level attention mechanism leverages emoji information. We define a clause as a short sentence separated by any punctuation with a sequence length (referring to the original sequence of words) greater than three. The basic idea of **TaneNet** is that multiple words form a clause and multiple clauses with emojis further combine to form a document.

As shown in Fig. 3, we first transform the high-dimensional word vectors into low-dimensional ones through the embedding layer. Then, we use a Bi-directional Gated Recurrent Units (BiGRU) network to extract the features of each word segment. Through the word-level attention mechanism, we obtain the vector representation of each clause, which we feed into another BiGRU layer to further extract characteristics of the clause segmentation. Next, we use a clause-level attention mechanism that incorporates the representation of emojis to obtain the document representation vector. Finally, we classify the document vector using a fully connected layer to obtain the final classification result.

We use Cross Entropy Loss as our loss function which could be written as:

$$L = \sum_{d \in D} \sum_{m=1}^{M} y_d \log(p_d), \quad (4)$$

where M represents the number of classification categories, D represents the size of the data set, y_d represents the true label of the dth document, and pd represents the predicted label of the dth document.

1) INPUT & EMBEDDING

The input is divided into two parts, one is the initial input of documents, and the other is the emojis representation that input at the clause level.

a: DOCUMENT

In order to adapt to the two-layer attention mechanism network, we should represent the input data as a three-dimensional matrix. According to the definition of the clause, we separate the document into multiple clauses, each clause is made up of words and expresses as $c_i : \{w_{i1}, w_{i2}, \ldots w_{iT}\}$, $(i \in [1, L])$, where w_{i1} represents the first word in clause c_i, the maximum length of clauses L and words T are given constants. If the separation result is less than the maximum length, the remaining value will be filled with zero. If the length is greater than the maximum length, the excess part will be discarded. Finally, the multiple clauses stacking up as the document $x \in R^{L \times T}$, that is a two-dimensional matrix. In Section III-A3, the sentiment lexicon has been constructed. We map the list to the two-dimensional matrix that just constructed, and complete the word embedding of the entire document $x \in R^{L \times J \times d}$, where d is the dimension of word embedding.

b: EMOJI

In order to maintain consistency between the dimensionality of the emoji vectors and the data in the clause-level attention

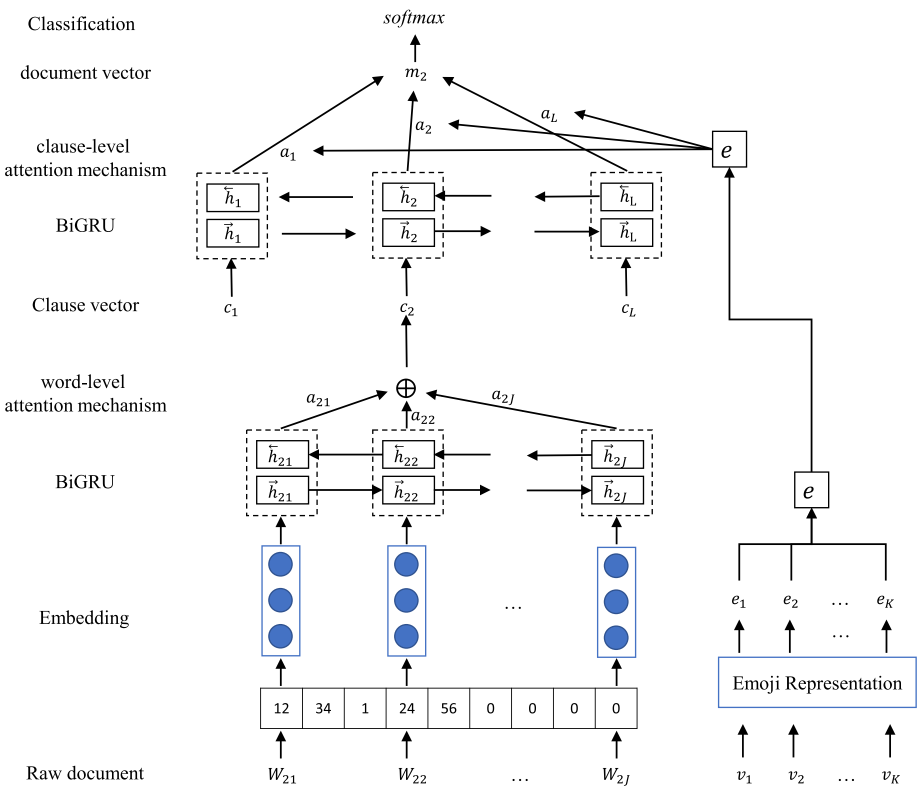


FIGURE 3. **The structure of TaneNet.**

mechanism, it is necessary to reduce the dimensionality of these vectors. Additionally, since a document may contain multiple instances of the same emoji, we employ a weighted summation method based on the frequency of appearance (k, including repetitions) to compute the representation v of the emojis in the document. The calculation formula is as follows:

$$\mathbf{v} = \frac{1}{k}\sum_{n=1}^{N} c_i \mathbf{e}_i, \tag{5}$$

where N represents the number of unique emojis in the document, c_i represents the count of occurrences of the i-th emoji, and e_i represents the vector representation of the i-th emoji. The formula computes the weighted average of the emojis based on their counts of appearance in the document, thereby obtaining the emojis' representation vector v for the document.

2) CLAUSE REPRESENTATION

For each clause c_i represented by word embeddings in the document, we first perform word encoding using a Bi-directional Gated Recurrent Units (BiGRU) network to extract the features of each word segment. Then, we use a word-level attention mechanism to obtain the vector representation of the clause.

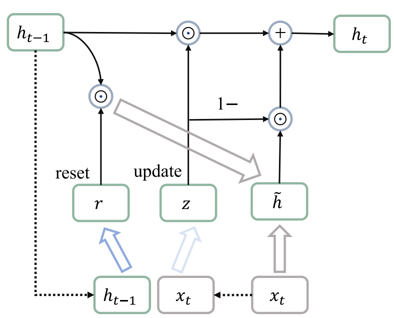


FIGURE 4. **The structure of gated recurrent units.**

a: WORD ENCODING

We employ a Bi-directional Gated Recurrent Units (BiGRU) neural network to extract the temporal features of words in each clause. The BiGRU can capture both past and future information of the current time step, by combining two GRUs with reversed temporal direction. The architecture of a single GRU block is shown in Fig. 4.

For a given input vector x_t at time step t, the calculating process is as follows:

$$h_t = (1 - z_t) \odot \widetilde{h_t} + z_t \odot h_{t-1}, \tag{6}$$

$$z_t = \sigma(W_{xz}x_t + W_{hz}h_{t-1} + b_z), \tag{7}$$

$$\widetilde{h_t} = \tanh(W_{xh}x_t + r_t \odot W_{hh}h_{t-1} + b_h), \tag{8}$$

$$r_t = \sigma(W_{xr}x_t + W_{hr}h_{t-1} + b_r), \tag{9}$$

where h_{t-1} denotes the previous hidden state, the candidate hidden state is $\widetilde{h}_t$. z_t denotes the update gate at time step t, decides how the information will be updated, depends on x_t and h_{t-1}. σ means the logistic sigmoid function, W, b are the weight and bias parameters of GRU that will be learned during training period. r_t denotes the reset gate at time step t, it decides the portion of past information devotes to $\widetilde{h}_t$.

We put the initial word representation x_{it} into BiGRU, and the output can be formulated as

$$h_{it} = [\overrightarrow{h_{it}}, \overleftarrow{h_{it}}], \tag{10}$$

$$\overrightarrow{h_{it}} = \overrightarrow{\mathrm{GRU}}(x_{it}),\ \ t \in [1, T], \tag{11}$$

$$\overleftarrow{h_{it}} = \overleftarrow{\mathrm{GRU}}(x_{it}),\ \ t \in [T, 1]. \tag{12}$$

Each word w_{it} can be represented as a combination of the hidden state of the forward and the reverse GRU network. This vector h_{it} describes the information near word w_{it} (including before and after).

b: WORD ATTENTION MECHANISM

To extract the most significant word features that contribute to the overall emotion of the entire clause, we employ the word-level attention mechanism, which assigns greater weight to important words. In this method, the alignment score quantifies the level of "Attention" given by the decoder to each output of the encoder when generating the subsequent output. It can be interpreted as the importance score of each word in each clause. To describe the alignment score, we utilize the additive attention mechanism, expressed as follows:

$$u_{ij} = V_w^T \tanh(W_w h_{ij} + b_w), \tag{13}$$

$$\alpha_{ij} = \frac{\exp(u_{ij})}{\sum_j \exp(u_{ij})}, \tag{14}$$

$$c_i = \sum_j \alpha_{ij} h_{ij}, \tag{15}$$

where h_{ij} is the word vector. Then the alignment score u_{ij} normalized by (14), the obtained α_{ij} is used as the weight of word vector h_{ij}. We use (15) to get the weighted clause vector c_i. At this point, the word vectors of different importance can be reflected in the obtained clause vector.

3) DOCUMENT REPRESENTATION

For each clause vector c_i in the document, we perform clause encoding and clause-level attention fusion with emoji, and finally obtain the representation vector of the entire document.

a: CLAUSE ENCODING

Similar to the word encoding, we put the sequence of clause representation $c_i(i \in [1, L])$ into BiGRU which combines temporal information in clause vector. The calculating process is as follows:

$$\overrightarrow{h_i} = \overrightarrow{\mathrm{GRU}}(c_i),\ \ i \in [1, L], \tag{16}$$

TABLE 3. The statistical information of lexicons.

DATASETS	#POS	#NEG	#TOTAL	WORDS
SELF-BUILT DATASET	8528	2954	11482	29777
NLPCC2013	6248	6250	12498	37842

$$\overleftarrow{h_i} = \overleftarrow{\mathrm{GRU}}(c_i),\ \ i \in [L, 1]. \tag{17}$$

By splicing the hidden layer states of the current clause c_i, we obtain the vector h_i include the information near the clause c_i.

b: CLAUSE ATTENTION MECHANISM

Since emojis have strong emotional representation information, we take into consideration not only the contribution of different clauses, but also the influence of emojis on clauses. This is achieved by incorporating the emoji representation vector e into the class-level attention mechanism. The calculation formula is as follows:

$$\mathrm{score}(h_i, e) = V_c^T \tanh(W_c h_i + W_e e + b_c), \tag{18}$$

$$\alpha_i = \frac{\exp(\mathrm{score}(h_i, e))}{\sum_i \exp(\mathrm{score}(h_i, e))}, \tag{19}$$

$$s_i = \sum_i \alpha_i c_i, \tag{20}$$

where $score(h_i, e)$ indicates that the importance score of each clause is a function of h_i and e. With the weights α_i under the influence of clauses and emoji, we calculate the weighted sum and finally get the document representation vector s_i.

4) SENTIMENT CLASSIFICATION

In the final layer, we input the document representation vector s_i into the softmax function for classification:

$$p = \mathrm{softmax}(W_s s + b_s), \tag{21}$$

and the class with the highest probability in the output result is the classification result. Then we use the Adadelta [43] as optimizer.

IV. EXPERIMENTS

A. EXPERIMENTAL SETTINGS

In order to verify the effectiveness of the method proposed, we choose to conduct experiments on two datasets which are suitable for emotional binary classification problems. One is the Weibo dataset that we built before, the other is NLPCC2013 which is widely used for sentiment analysis of Chinese Weibo. The statistical information of these lexicons are exhibited in Table 3.

The aforementioned table highlights certain imbalances in our self-built dataset. Specifically, we note that the average number of clauses in each Weibo document is 4, the maximum number of clauses is 25, and the average number of words in each document is 37.

To ensure the stability of experimental results, we choose 5-folds cross-validation and split the train and test set of data

TABLE 4. The statistical of three subsets.

Subsets	Followers number	#User	#Neg	#Pos	#Total	#T	#U
Data1	<2000	207	1817	3358	5175	49.53%	89.86%
Data2	2000∼100,000	181	632	2300	2932	56.17%	87.29%
Data3	>100,000	192	505	2870	3375	66.70%	92.71%

by 8 : 2. The word embedding vectors are 300 dimensions. In the attention mechanism, the weights are initialized with the glorot distribution. We use ModelCheckpoint and EarlyStopping to accelerate training, use $L2$ regularization and DropOut to solve the overfitting problem.

Besides, we use classification accuracy as the evaluation metric which can be formulated as $\text{Accuracy} = T/N$, where T and N denotes the number of documents correctly classified and test set respectively.

B. EXPERIMENTAL RESULTS

1) PERFORMANCE OF EMOJI REPRESENTATIONS

Here we verify the validity of the emojis' sentiment information obtained in the multi-dimensional emoji representation. We add emojis to the basic sentiment dictionary of this article so that form a new sentiment dictionary, and compare it with other unsupervised sentiment analysis methods. We select two common unsupervised sentiment analysis methods: JST [44] and SLDA [45].

Joint Sentiment-Topic (JST): a method of joint emotion-topic model based on LDA (Latent Dirichlet Allocation), it can detect emotions and topics from the text at the same time, which improves the defect that topics and emotions are separately detected in traditional methods. It is easy to transfer to other fields.

Sentiment-LDA (SLDA): SLDA considers not only the global theme of the entire document but also the local associations between words. It noticed that the emotion of words depends on the local context, so it relaxes the assumption of emotion dependence, treats the emotion of words as a Markov chain.

To comprehensively analyze the impact of social media influence on emoji usage in sentiment analysis, we divided our self-built dataset into three subsets based on the number of followers. This stratification allows us to assess how different levels of user engagement and visibility might influence the use and interpretation of emojis in sentiment expression. The subsets are defined as follows: users with fewer than 2,000 followers, users with 2,000 to 100,000 followers, and users with more than 100,000 followers.

As shown in Table 4, the distribution of emoji usage varies significantly across these subsets. The column #T, representing the proportion of microblogs using emojis, and #U, indicating the proportion of users who employ emojis, highlight distinct patterns of emoji utilization among different user groups. For instance, users with a larger following tend to use emojis more consistently, which could be attributed to their need to maintain clear and effective communication with a broad audience. Conversely, users with fewer followers might use emojis less frequently or in more nuanced ways, potentially reflecting more personal or less public communication styles.

TABLE 5. Comparison of various methods on accuracy.

Model	Data1	Data2	Data3	AVG
Basic Sentiment Dictionary	<u>59.98</u>	55.05	57.07	57.367
JST	54.11	44.99	42.22	47.107
Sentiment-LDA	58.53	<u>66.58</u>	<u>68.86</u>	<u>64.657</u>
Emoji-based Sentiment Dictionary	**69.57**	**67.68**	**75.50**	**70.917**

We conduct experiments on three data subsets separately and the results obtained are shown in Table 5. We can draw the following conclusions: 1) The accuracy of the emoji-based dictionary method on the three data subsets is significantly higher than that of JST and SLDA; 2) Emojis have a strong emotional representation; 3)The accuracy of the emoji-based dictionary method is positively correlated with the proportion of Weibo containing emojis, but has little relation with the proportion of users.

2) PERFORMANCE OF TANENET

To verify the effectiveness of the proposed sentiment analysis model, We use not only traditional machine learning methods such as Support Vector Machine (SVM), Naïve Bayes (NB) and Logistic Regression (LR) but also deep learning methods such as original GRU, LSTM and BiGRU as baseline models. Besides, we also discuss the differences between bag of words and word embedding. The experimental results are shown in Table 6, where Bow denotes bag of words, SG denotes Skip-gram.

We also compare the latest approaches, namely BERT [46], RoBERTa [47] and ALBERT-FAET [48], where BERT is a bi-directional encoder model that uses the transformer encoder block for linking. BERT directly utilizes the Encoder module in the Transformer architecture and discards the Decoder module, which automatically has bi-directional encoding capability and powerful feature extraction capability. Text embedding with the BERT model has yielded very good results. RoBERTa is a variant of the BERT model, which changes the training data from 16G to 160G, making the training time of RoBERTa longer and the training model larger. In addition, the dynamical mask mechanism is adjusted, and the experimental results show that RoBERTa is more effective. The ALBERT-FAET approach obtains text embeddings by using an ALBERT pre-training

TABLE 6. The results of various classifiers on accuracy.

Model	Fold-1	Fold-2	Fold-3	Fold-4	Fold-5	AVG
Bow+SVM	79.45	78.1	79.71	79.97	79.74	79.39
Bow+NB	82.19	81.58	82.67	83.01	83.09	82.51
SG+SVM	83.80	83.67	83.54	84.93	84.53	84.10
SG+LR	83.89	83.46	83.63	84.67	84.27	83.98
SG+BiGRU	85.07	85.41	85.02	84.84	86.54	85.18
SG+BiLSTM	85.90	85.07	85.72	85.76	87.80	85.64
RoBERTa	85.79	87.48	87.75	87.11	85.71	86.87
BERT	85.24	83.09	86.61	84.19	85.96	85.63
ALBERT-FAET	84.43	85.52	86.21	85.01	85.43	85.32
MA-BiGSNP	86.73	87.83	87.94	88.19	85.19	87.18
ELSA	88.42	87.78	88.21	88.38	88.28	88.21
TaneNet	**88.90**	**89.07**	**88.72**	**89.76**	**90.90**	**89.45**

model and learns emoji embeddings using an attention-based LSTM, where ALBERT is also a variant of BERT. Besides, a fine-grained attention mechanism is proposed to capture word-level interactions between text only and emojis. Finally, we connect these features and feed them into a CNN classifier to predict the sentiment labels of microblogs.

TaneNet's performance was also assessed in the context of cross-lingual sentiment analysis using ELSA [49] (Emoji-driven Language Sentiment Analysis). ELSA utilizes emoji prediction to learn sentiment patterns across languages, offering an innovative method to bridge linguistic gaps without traditional translation, thus enhancing sentiment analysis in multilingual contexts.

Additionally, the study includes a comparison with MA-GSNP [50], a novel model proposed in 2024 that incorporates a bidirectional GSNP (BiGSNP) combined with a multi-attention mechanism. This approach captures semantic correlations and emphasizes word significance more effectively, offering advanced capabilities in understanding complex sentence structures.In summary, we compare the above-mentioned methods and the results is shown in Table 6. In order to see the performance of each algorithm more intuitively, Fig. 5 better shows the results of each algorithm.

Results from Table 6 show that SG+SVM outperforms BoW+SVM, highlighting the efficacy of word embeddings over sparse vector representations. Among the deep learning approaches, GRU consistently demonstrated improved performance over the traditional models, underscoring the advantage of capturing temporal dependencies in text data.

The introduction of emojis through the ELSA model and the incorporation of two-level sequence processing in the MA-GSNP model are critical innovations that significantly enhance sentiment analysis. ELSA utilizes emojis to provide additional sentiment context, improving accuracy across different languages. In contrast, MA-GSNP employs a two-level sequence approach that captures deeper semantic correlations both at the word and clause levels, effectively enriching the model's interpretative depth.

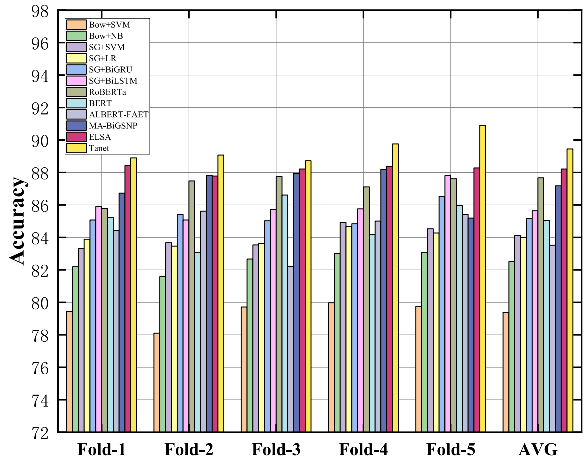


FIGURE 5. Comparison of various classifiers on accuracy.

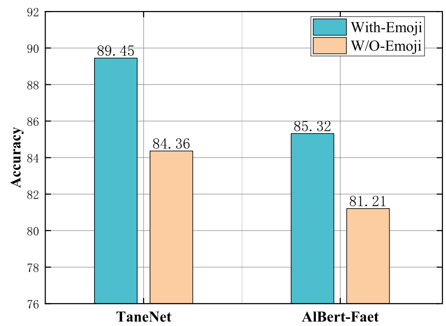


FIGURE 6. Ablation experiment results:Emoji Impact.

TaneNet, which integrates these advanced strategies, consistently outperforms existing methods. Its comprehensive framework leverages both emoji-enhanced representations and detailed sequence analysis, resulting in superior sentiment classification accuracy. This demonstrates TaneNet's effectiveness in handling complex sentiment analysis challenges, setting a new standard in the field.

C. ABLATION STUDY

It is essential to explore the impact of emojis on sentiment polarity in a more comprehensive manner. To achieve this, we adopted a rigorous approach by selecting the top three advanced methods that consider emojis and conducted ablation experiments. By removing the emoji factor and analyzing the algorithms' performance, we obtained a clearer understanding of the effect of emoji on sentiment analysis. In this regard, we performed experiments using state-of-the-art models, including BERT [46] and AlBERT-FAET [48], along with the method proposed in this paper. The experimental results are presented in Fig. 6, and it is evident that the accuracy of each method significantly improved after integrating the emoji factor. The inclusion of emoji as a crucial factor in sentiment analysis is crucial, and our study contributes to improving the accuracy of sentiment analysis by considering the influence of emoji on sentence polarity.

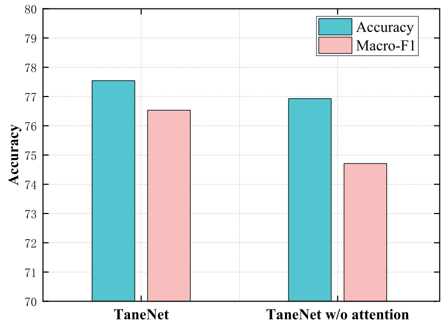


FIGURE 7. Ablation experiment results:Attention Impact.

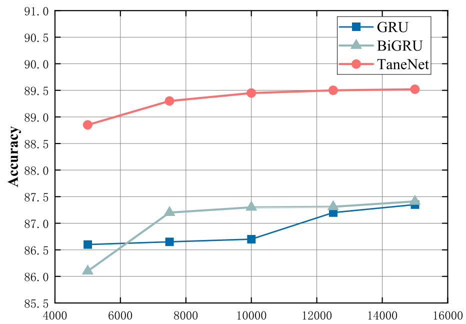


FIGURE 8. Experiment of lexicon size on self-built data set.

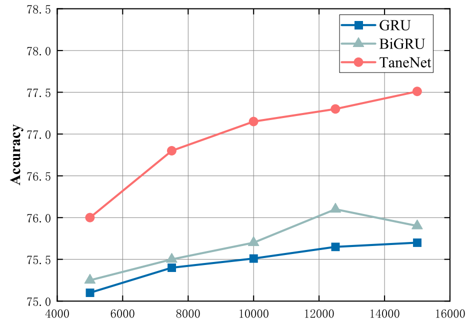


FIGURE 9. Experiment of lexicon size on NLPCC.

TABLE 7. The running time in self-built dataset.

Model	5000	7500	10000	12500	15000
GRU	565	586	576	768	736
BiGRU	1616	1733	1657	1837	1846
TaneNet	**443**	**439**	**447**	**539**	**617**

TABLE 8. The running time in NLPCC.

Model	5000	7500	10000	12500	15000
GRU	923	771	955	1152	1062
BiGRU	1690	1751	1838	1899	1905
TaneNet	**468**	**547**	**539**	**586**	**702**

In order to verify the performance of the Word and Clause attention mechanism in our proposed TaneNet model, we performed ablation experiments on the published NLPCC dataset, in which Tanenet without attention represents that we removed the word-level attention mechanism and clause attention mechanism, so the model cannot effectively focus on the key parts of the document. As shown in Fig. 7, we can observe that the removal attention mechanism module has a slight downward trend compared with the TaneNet model, which shows that the attention mechanism module can extract the key information of documents, strengthen the learning representation of documents, and effectively improve the performance of document sentiment analysis.

D. EXPERIMENTAL ANALYSIS

1) THE SIZE OF SENTIMENT LEXICON

We experimented with different sizes on the two data sets. Here, we choose the top 5, 000 to 15, 000 words in terms of frequency to form a lexicon for experiment. As we can see from Fig. 8 and Fig. 9, when the size of the lexicon increases, the accuracy of all models shows an overall increasing trend on the two datasets. At first, the growth rate is very fast, and then the growth rate gradually slows down. This shows that after removing the stop words, the more frequent words, the greater the contribution to classification.

Additionally, in order to measure the impact of the lexicon size on time and accuracy, we also record the running time of the program under different vocabulary sizes. From Table 7 and Table 8, it is easily found that as the lexicon size grows, the running time of the model continues to increase. Because BiGRU is composed of two GRUs running in opposite directions, its running time is about twice that of GRU. Although our model uses BiGRU, the running time of the TaneNet model only accounts for about 1/3 of BiGRU, which shows that the TaneNet model can converge faster. As the lexicon size increases, the model's performance generally improves, as there are more words for the model to learn from. However, this improvement in performance may eventually plateau or even decrease as the size of the lexicon becomes too large, causing the model to struggle with effectively learning from the additional words. Therefore, it is important to find a balance between lexicon size and model performance. By measuring the model's time and accuracy under different vocabulary sizes, we can determine the optimal lexicon size for our specific application. This information can then be used to optimize optimize the model's performance and improve its efficiency.

2) THE SELECTION OF RECURRENT NEURAL NETWORKS

ALBERT-FAET and TaneNet have similarities and differences, where ALBERT-FAET used the attention network-based BiLSTM, we used the attention network-based BiGRU neural network. To better compaer the two methods above, we conducted experiments with BiLSTM and BiGRU respectively, and the experimental results are shown in Fig. 10, It is can be concluded that BiGRU is more advantageous.

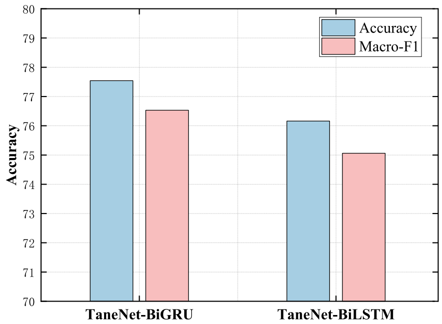


FIGURE 10. The selection experiment of recurrent neural networks.

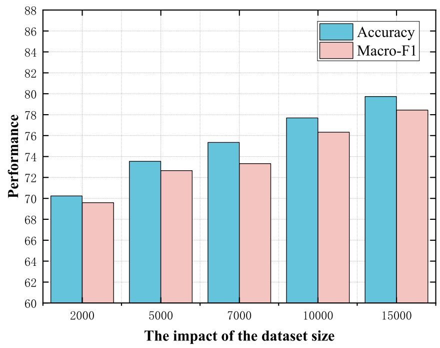


FIGURE 11. The impact of the dataset size on TaneNet.

3) THE IMPACT OF THE DATASET SIZE

In order to verify the impact of different dataset sizes on our proposed TaneNet model, we divide the published NLPCC dataset into different subsets. We set the size of subset 1 to be 2000, that of subset 2 to be 5000, that of subset 3 to be 7000, that of subset 4 to be 10000 and that of subset 5 to be 15000. We can observe in Fig. 11, that with the increase of dataset size, the performance of the model is gradually improved, which shows that a larger dataset can better train data and improve the performance of model classification.

4) THE IMPACT OF THE SCALE OF EXPLICIT EMOJIS

In this section, we analyze the impact of the scale of Explicit Emojis in our proposed model TaneNet. In order to verify the impact of Explicit Emojis with different scales, we set 20, 40, 60, 80, 100 different scales to conduct experiments respectively, and the experimental results are shown in Fig. 12. we can find that Explicit Emojis with the scale of 40 achieves the best performance, which is better than other Explicit Emojis scales. Therefore, we set the scale of Explicit Emojis as 40. When the scale of Explicit Emojis is less than 40, the model does not achieve the best effect, which shows that fewer emojis are not enough to fully express the emotional information of the text. When the scale of Explicit Emojis is greater than 40, the performance of the model is significantly reduced, which may be because more emoji will increase the sparsity of data, which is not conducive to the mining of emotional information.

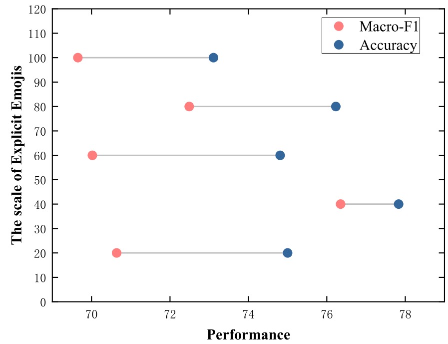


FIGURE 12. The impact of the scale of Explicit Emojis.

V. CONCLUSION

This paper introduces a novel approach to sentiment analysis by incorporating emojis and user information to reduce ambiguity in short text, such as microblogs. Our method innovatively models the characteristics, position, and usage frequency of emojis and integrates this representation into a two-level attention mechanism. This mechanism, which includes both word-level and clause-level attention, is particularly effective for analyzing the compact and often nuanced content of microblogs. By focusing on the emotional impact of emojis within clauses, our approach not only identifies key expressive elements within the text but also significantly enhances the accuracy of sentiment analysis.

The empirical results demonstrate that our model, TaneNet, outperforms existing state-of-the-art methods, particularly in contexts where traditional text-based analysis methods fall short. By effectively leveraging emoji-enhanced representations, TaneNet provides a more nuanced understanding of sentiment, suggesting that emojis are not merely decorative but play a substantive role in communication.

Nevertheless, there are certain aspects that warrant further investigation. The context in which emojis are used, for instance, presents an intricate area for deeper analysis. Moreover, the semantic mining of emojis, especially in contexts involving sarcasm or irony, introduces complexities where the emoji's apparent sentiment may diverge from its intended meaning. These considerations have been thoughtfully reflected in the revised discussion of limitations, enhancing the manuscript by providing a balanced view of our study's scope and pointing to avenues for future research.

Looking ahead, our future studies will explore several promising directions:

1) Model Enhancement: We plan to evaluate other word embedding models, such as GloVe and FastText, to determine the most optimal tool for integrating with the TaneNet framework. This exploration will help refine our approach by potentially capturing a broader array of semantic nuances.
2) Geographical and Cultural Expansion: Although our current study focuses on the Chinese context, the universal nature of emojis suggests that TaneNet could

be adapted for multilingual settings. Future research will aim to test and optimize TaneNet across different languages and cultural contexts, which could broaden its applicability and effectiveness in global sentiment analysis.

3) Deeper Semantic Mining: We also intend to delve deeper into the semantic mining of emojis, especially to better handle cases of sarcasm and irony where the traditional sentiment polarity of an emoji may not align with its intended usage in communication.

REFERENCES

[1] W. Huang, Z. Li, L. Zhang, and Y. Li, "Review of intelligent microblog short text processing," *Web Intell.*, vol. 14, no. 3, pp. 211–228, Aug. 2016.

[2] L. Wang, J. Niu, and S. Yu, "SentiDiff: Combining textual information and sentiment diffusion patterns for Twitter sentiment analysis," *IEEE Trans. Knowl. Data Eng.*, vol. 32, no. 10, pp. 2026–2039, Oct. 2020.

[3] S. Wijeratne, L. Balasuriya, A. Sheth, and D. Doran, "EmojiNet: Building a machine readable sense inventory for emoji," in *Social Informatics*. Cham, Switzerland: Springer, 2016, pp. 527–541.

[4] A. Hogenboom, D. Bal, F. Frasincar, M. Bal, F. de Jong, and U. Kaymak, "Exploiting emoticons in sentiment analysis," in *Proc. 28th Annu. ACM Symp. Appl. Comput.*, Mar. 2013, pp. 703–710.

[5] S. Gupta, A. Singh, and J. Ranjan, "Sentiment analysis: Usage of text and emoji for expressing sentiments," in *Advances in Data and Information Sciences*. Singapore: Springer, 2020, pp. 477–486.

[6] M. S. Amrullah, I. Budi, A. B. Santoso, and P. K. Putra, "The effect of using emoji and hashtag in sentiment analysis on Twitter case study: Indonesian online travel agent," *AIP Conf. Proc.*, vol. 2654, no. 1, 2023, Art. no. 020013.

[7] S. A. A. Hakami, R. Hendley, and P. Smith, "Gender impact on emoji sentiment analysis among Arabic users in digital networks," in *Proc. 14th Int. Conf. Comput. Intell. Commun. Netw. (CICN)*, Dec. 2022, pp. 199–203.

[8] Q. Bai, Q. Dan, Z. Mu, and M. Yang, "A systematic review of emoji: Current research and future perspectives," *Frontiers Psychol.*, vol. 10, p. 2221, Oct. 2019.

[9] Y. Wang, R. Li, H. Zhang, H. Tan, and Q. Chai, "Using sentence-level neural network models for multiple-choice reading comprehension tasks," *Wireless Commun. Mobile Comput.*, vol. 2018, pp. 1–8, Jul. 2018.

[10] S. Huang, Q. Zhao, X.-Z. Xu, B. Zhang, and D. Wang, "Emojis-based recurrent neural network for Chinese microblogs sentiment analysis," in *Proc. IEEE Int. Conf. Service Operations Logistics, Informat. (SOLI)*, Nov. 2019, pp. 59–64.

[11] F. Huang, S. Zhang, J. Zhang, and G. Yu, "Multimodal learning for topic sentiment analysis in microblogging," *Neurocomputing*, vol. 253, pp. 144–153, Aug. 2017.

[12] F. H. Khan, U. Qamar, and S. Bashir, "A semi-supervised approach to sentiment analysis using revised sentiment strength based on SentiWordNet," *Knowl. Inf. Syst.*, vol. 51, no. 3, pp. 851–872, Jun. 2017.

[13] S. Baccianella, A. Esuli, and F. Sebastiani, "SENTIWORDNET 3.0: An enhanced lexical resource for sentiment analysis and opinion mining," in *Proc. 7th Int. Conf. Lang. Resour. Eval. (LREC)*, vol. 10, 2010, pp. 2200–2204.

[14] B. Zhang, D. Xu, H. Zhang, and M. Li, "STCS lexicon: Spectral-clustering-based topic-specific Chinese sentiment lexicon construction for social networks," *IEEE Trans. Computat. Social Syst.*, vol. 6, no. 6, pp. 1180–1189, Dec. 2019.

[15] U. Nuha and C.-H. Lin, "Aspect-based sentiment analysis with semi-supervised approach on Taiwan social distancing app user reviews," in *Proc. Int. Conf. Artif. Intell. Inf. Commun. (ICAIIC)*, Feb. 2023, pp. 444–447.

[16] L. G. Singh, A. Anil, and S. R. Singh, "SHE: Sentiment hashtag embedding through multitask learning," *IEEE Trans. Computat. Social Syst.*, vol. 7, no. 2, pp. 417–424, Apr. 2020.

[17] H. Peng, Y. Ma, Y. Li, and E. Cambria, "Learning multi-grained aspect target sequence for Chinese sentiment analysis," *Knowl.-Based Syst.*, vol. 148, pp. 167–176, May 2018.

[18] B. Pang, L. Lee, and S. Vaithyanathan, "Thumbs up? Sentiment classification using machine learning techniques," 2002, *arXiv:cs/0205070*.

[19] G. Rao, W. Huang, Z. Feng, and Q. Cong, "LSTM with sentence representations for document-level sentiment classification," *Neurocomputing*, vol. 308, pp. 49–57, Sep. 2018.

[20] G. Alexandridis, J. Aliprantis, K. Michalakis, K. Korovesis, P. Tsantilas, and G. Caridakis, "A knowledge-based deep learning architecture for aspect-based sentiment analysis," *Int. J. Neural Syst.*, vol. 31, no. 10, Oct. 2021, Art. no. 2150046.

[21] P. Routray, C. K. Swain, and R. C. Balabantaray, "Sentiment analysis from student feedbacks using supervised machine learning approaches," in *Innovations in Intelligent Computing and Communication*. Cham, Switzerland: Springer, 2022, pp. 273–288.

[22] G. V. Singh, M. Firdaus, A. Ekbal, and P. Bhattacharyya, "Unity in diversity: Multilabel emoji identification in tweets," *IEEE Trans. Computat. Social Syst.*, vol. 10, no. 3, pp. 1029–1038, 2022.

[23] S. Al-Azani and E.-S. El-Alfy, "Emojis-based sentiment classification of Arabic microblogs using deep recurrent neural networks," in *Proc. Int. Conf. Comput. Sci. Eng. (ICCSE)*, Mar. 2018, pp. 1–6.

[24] S. A. A. Hakami, R. Hendley, and P. Smith, "Emoji sentiment roles for sentiment analysis: A case study in Arabic texts," in *Proc. The 7th Arabic Natural Lang. Process. Workshop (WANLP)*, 2022, pp. 346–355.

[25] J. Chen, Z. Yao, S. Zhao, and Y. Zhang, "Fusion pre-trained emoji feature enhancement for sentiment analysis," *ACM Trans. Asian Low-Resource Lang. Inf. Process.*, vol. 22, no. 4, pp. 1–14, Apr. 2023.

[26] D. Li, R. Rzepka, M. Ptaszynski, and K. Araki, "Emoticon-aware recurrent neural network model for Chinese sentiment analysis," in *Proc. 9th Int. Conf. Awareness Sci. Technol. (iCAST)*, Sep. 2018, pp. 161–166.

[27] Q. Zhao, J. Huang, G. Liu, Y. Miao, and P. Wang, "A multiinterest and social interest-field framework for financial security," *IEEE Trans. Computat. Social Syst.*, vol. 11, no. 2, pp. 1685–1695, 2024.

[28] W. Yin, K. Kann, M. Yu, and H. Schütze, "Comparative study of CNN and RNN for natural language processing," 2017, *arXiv:1702.01923*.

[29] M. Ling, Q. Chen, Q. Sun, and Y. Jia, "Hybrid neural network for Sina Weibo sentiment analysis," *IEEE Trans. Computat. Social Syst.*, vol. 7, no. 4, pp. 983–990, Aug. 2020.

[30] W. Li, W. Shao, S. Ji, and E. Cambria, "BiERU: Bidirectional emotional recurrent unit for conversational sentiment analysis," *Neurocomputing*, vol. 467, pp. 73–82, Jan. 2022.

[31] Z. Wang, G. Xu, X. Zhou, J. Y. Kim, H. Zhu, and L. Deng, "Deep tensor evidence fusion network for sentiment classification," *IEEE Trans. Computat. Social Syst.*, early access, 2022, doi: 10.1109/TCSS.2022.3197994.

[32] M. U. Salur and I. Aydin, "A novel hybrid deep learning model for sentiment classification," *IEEE Access*, vol. 8, pp. 58080–58093, 2020.

[33] F. Liu, J. Zheng, L. Zheng, and C. Chen, "Combining attention-based bidirectional gated recurrent neural network and two-dimensional convolutional neural network for document-level sentiment classification," *Neurocomputing*, vol. 371, pp. 39–50, Jan. 2020.

[34] K. Xu, J. Ba, R. Kiros, K. Cho, A. Courville, R. Salakhudinov, R. Zemel, and Y. Bengio, "Show, attend and tell: Neural image caption generation with visual attention," *Comput. Sci.*, vol. 2015, pp. 2048–2057, Feb. 2015.

[35] R. Cao, G. Liu, Y. Xie, and C. Jiang, "Two-level attention model of representation learning for fraud detection," *IEEE Trans. Computat. Social Syst.*, vol. 8, no. 6, pp. 1291–1301, Dec. 2021.

[36] Z. Yang, D. Yang, C. Dyer, X. He, A. Smola, and E. Hovy, "Hierarchical attention networks for document classification," in *Proc. Conf. North Amer. Chapter Assoc. Comput. Linguistics, Human Lang. Technol.*, 2016, pp. 1480–1489.

[37] Y. Xu, H. Gao, R. Li, Y. Jiang, S. Mumtaz, Z. Jiang, J. Fang, and L. Dong, "Adversarial learning-based sentiment analysis for socially implemented IoMT systems," *IEEE Trans. Computat. Social Syst.*, vol. 10, no. 4, pp. 1691–1700, 2023.

[38] B. A. Chandio, A. S. Imran, M. Bakhtyar, S. M. Daudpota, and J. Baber, "Attention-based RU-BiLSTM sentiment analysis model for Roman Urdu," *Appl. Sci.*, vol. 12, no. 7, p. 3641, Apr. 2022.

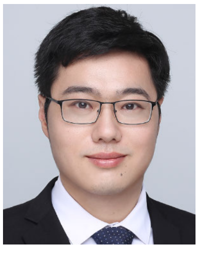
[39] H. Tan, S. Deng, T. Qian, and D. Ji, "Emoji-attentional neural network for microblog sentiment analysis," *Appl. Res. Comput.*, vol. 36, no. 9, pp. 2647–2650, 2019.

[40] C. Suman, S. Saha, and P. Bhattacharyya, "An attention-based multimodal Siamese architecture for tweet-user verification," *IEEE Trans. Computat. Social Syst.*, vol. 10, no. 5, pp. 2764–2772, 2023.

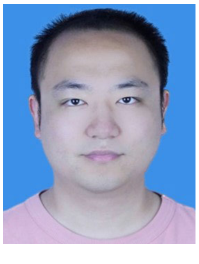
[41] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," 2013, *arXiv:1301.3781*.

[42] S. Lai, K. Liu, S. He, and J. Zhao, "How to generate a good word embedding," *IEEE Intell. Syst.*, vol. 31, no. 6, pp. 5–14, Nov. 2016.

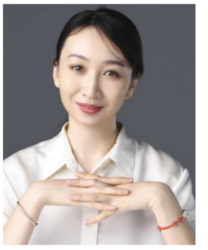
[43] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*.
[44] C. Lin, Y. He, R. Everson, and S. Ruger, "Weakly supervised joint sentiment-topic detection from text," *IEEE Trans. Knowl. Data Eng.*, vol. 24, no. 6, pp. 1134–1145, Jun. 2012.
[45] F. Li, M. Huang, and X. Zhu, "Sentiment analysis with global topics and local dependency," in *Proc. AAAI Conf. Artif. Intell.*, 2010, vol. 24, no. 1, pp. 1371–1376.
[46] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," 2018, *arXiv:1810.04805*.
[47] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, and V. Stoyanov, "RoBERTa: A robustly optimized BERT pretraining approach," 2019, *arXiv:1907.11692*.
[48] D. Yang, L. Kejian, Y. Cheng, F. Yuanyuan, and L. Weihao, "Emoji-based fine-grained attention network for sentiment analysis in the microblog comments," 2022, *arXiv:2206.12262*.
[49] Z. Chen, S. Shen, Z. Hu, X. Lu, Q. Mei, and X. Liu, "Emoji-powered representation learning for cross-lingual sentiment classification," in *Proc. World Wide Web Conf.*, May 2019, pp. 251–262.
[50] Y. Huang, X. Bai, Q. Liu, H. Peng, Q. Yang, and J. Wang, "Sentence-level sentiment classification based on multi-attention bidirectional gated spiking neural P systems," *Appl. Soft Comput.*, vol. 152, Feb. 2024, Art. no. 111231.



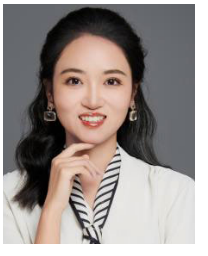
QIN ZHAO (Senior Member, IEEE) received the Ph.D. degree from the Department of Computer Science and Technology, Tongji University, Shanghai, China, in 2016. Currently, he is an Associate Professor with the Department of Computer Science and Technology, Shanghai Normal University, Shanghai. He is also the Deputy Director of Shanghai Engineering Research Center of Intelligent Education and Big Data. His research interests include new-generation artificial intelligence technologies, social network analysis, data mining, and intelligence in scientific computing. He has published more than 30 papers in premier international journals and conferences, including IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS: SYSTEMS, and IEEE TRANSACTIONS ON COMPUTATIONAL SOCIAL SYSTEMS. He has been recognized as a Senior Member of China Computer Federation (CCF).



PEIHAN WU received the B.E. degree from the Department of Computer Science and Technology, Wuhan University, in 2020. He is currently pursuing the M.S. degree with the Department of Computer Science and Technology, Shanghai Normal University, China. His research interests include community detection, natural language processing, and sentiment analysis.



JIE LIAN received the Ph.D. degree from Towson University, in 2017. She is currently an Associate Professor with the Department of Computer Science and Technology, Shanghai Normal University, where she has been a Faculty Member, since 2017. Her research interests include spatiotemporal data mining, deep learning, and big data, ranging from theory to design to implementation. She received Shanghai Sailing Talent Program, in 2019.



DONGDONG AN received the B.S. and Ph.D. degrees in software engineering from East China Normal University, Shanghai, China, in 2013 and 2020, respectively. She is currently a Lecturer with the Department of Computer Science and Technology, Shanghai Normal University, Shanghai. Her current research interests include model-driven architecture, machine learning, formal methods, and statistical model-checking techniques.

MAOZHEN LI received the Ph.D. degree from the Institute of Software, Chinese Academy of Sciences, Beijing, China, in 1997. He finished his postdoctoral research with the School of Computer Science and Informatics, Cardiff University, Cardiff, U.K., from 1999 to 2002. He is currently a Professor with the Department of Electronic and Computer Engineering, Brunel University London, London, U.K. His research interests include high-performance computing including cloud computing and edge computing, big data analytics, and intelligent systems, with an application on smart grids and smart cities. He has more than 200 research publications in these areas, including four books. He is a fellow of the British Computer Society (BCS) and the Institute of Engineering and Technology (IET). He has served more than 30 IEEE conferences and serves on the editorial board for a number of journals.