

Received June 26, 2020, accepted July 12, 2020, date of publication July 14, 2020, date of current version July 6, 2021.

Digital Object Identifier 10.1109/ACCESS.2020.3009292

A Novel Emotion Lexicon for Chinese Emotional Expression Analysis on Weibo: Using Grounded Theory and Semi-Automatic Methods

LIANG XU¹, LINJIAN LI¹, ZEHUA JIANG¹, ZAOYI SUN², XIN WEN¹,
JIAMING SHI¹, RUI SUN¹, AND XIUYING QIAN¹

¹Department of Psychology and Behavioral Science, Zhejiang University, Hangzhou 310058, China

²College of Education, Zhejiang University of Technology, Hangzhou 310014, China

Corresponding author: Xiuying Qian (xqian@zju.edu.cn)

ABSTRACT As one of the most popular social media platforms in China, Weibo has aggregated huge numbers of texts containing people's thoughts, feelings, and experiences. Analyzing emotions expressed on Weibo has attracted a great deal of academic attention. Emotion lexicon is a vital foundation of sentiment analysis, but the existing lexicons still have defects such as a limited variety of emotions, poor cross-scenario adaptability, and confusing written and online expressions and words. By combining grounded theory and semi-automatic methods, we built a Weibo-based emotion lexicon for sentiment analysis. We first took a bottom-up approach to derive a theoretical model for emotions expressed on Weibo, and the substantive coding led to eight core emotion categories: joy, expectation, love, anger, anxiety, disgust, sadness, and surprise. Second, we built a new emotion lexicon containing 2,964 words by manually selecting seed words, constructing a word vector model to expand words, and making rules to filter words. Finally, we tested the effectiveness of our lexicon by using a lexicon-based approach to recognize the emotions expressed in Weibo text. The results showed that our lexicon performed better in Weibo emotion recognition than five other Chinese emotion lexicons. This study proposed a method to construct an emotion lexicon that considered both theory and application by combining qualitative research and artificial intelligence methods. Our work also provided a reference for future research in the field of social media sentiment analysis.

INDEX TERMS Emotion lexicon, emotional expression analysis, grounded theory, Weibo.

I. INTRODUCTION

Many social media platforms encourage emotional self-expression, inviting users to regularly update their opinions on their personal life, social events, and so forth [1]. As one of China's most popular social platforms, Weibo has aggregated huge numbers of texts that contain people's thoughts, feelings, and experiences. And analyzing emotions expressed on Weibo has attracted much academic attention [2]–[4]. In the past two decades, computer scientists have expended a lot of effort on sentiment analysis, a research area that extracts emotions from natural language texts [5]. Their work has made this technology accessible for researchers in other fields and has been widely applied to depression diagnosis [6], opinion mining [7] and financial prediction [8].

Lexicon-based approaches and machine learning based approaches are two common approaches for sentiment analysis [9]. The lexicon-based method uses the emotion lexicon,

a list of words and expressions used to express people's emotions to label the emotion words in a document [10]. Therefore, building a high-quality emotion lexicon is crucial. Previous studies have created different sentiment or emotion lexicons in the Chinese language, such as the Simplified Chinese-Linguistic Inquiry and Word Count (SC-LIWC) [11], the National Taiwan University Sentiment Dictionary (NTUSD) [12], HowNet [13], and the Tsinghua Open Chinese Lexicon (THUOCL) [14]. But these lexicons still have shortcomings. First, most lexicons are only classified based on the polarity (positive or negative), ignoring the complexity of emotions. Second, even if discrete emotions have been considered (e.g., NTUSD), the choice of emotion categories is subjective and lacks theoretical support. Third, emotional expression and description words are often confused, and the words used to describe personal emotions in writing are not entirely applicable to emotional expressions on social networks. Therefore, it is necessary to build a more granular and theoretically supported lexicon for analyzing emotional expressions on social media.

The associate editor coordinating the review of this manuscript and approving it for publication was Shiqing Zhang¹.

In addition, the work of [15] showed that the norms of expressing emotions on social media are different, which reminds us that emotion lexicons based on different social media platforms may also differ. Therefore, we decided to construct a Weibo-based emotion lexicon for Chinese sentiment analysis. To make the lexicon's emotion categories more appropriate, the current study took a bottom-up approach. We first built a theoretical model of Weibo emotion expression with associated seed words based on the grounded theory approach [16]. We then used a semi-automatic method to construct the emotion lexicon. Finally, the new emotion lexicon was used to recognize emotions expressed in Weibo text founded on a lexicon-based approach and then compared with other emotion lexicons.

II. LITERATURE REVIEW

A. EMOTION EXPRESSION ON SOCIAL MEDIA

As human beings, we are born with emotional expression. Expressing our emotions is rewarding intrinsically [17]. It improves our social relationships [18], [19] and benefits our well-being [20], [21]. With the development of technology, social media has become a platform where people can promote their opinions and feelings. Social media encourages more emotional expression, not only as a strategy but as a users' needs [1]. The controllability of self-presentation, reduced communication cues, and potentially asynchronous communication in social media elicit more frequent expression online [15], [22]. Such abundant expression turns social media into a huge resource for expression of emotion.

A series of derivative issues of emotional expression on social media have been widely discussed. For instance, previous researchers have shown that social media is the garden of emotional contagion. The emotional expression of others can affect users' own emotional expression on both Facebook [23] and Twitter [24]. This means that emotions can be transmitted by one to another through emotional expression on the social media. The work of [25] found that the expression of emotion is also key for the spread of moral and political ideas in online social networks. Emotional contagion outbursts on social media attract massive transmissions and may trigger historical events with the further spreading of fervent and intense emotions. Waterloo *et al.* focused their research on the norms of online expressions of emotion and found that the expression of emotions was rated more appropriately for a private platform than a public, social media platform [15]. In addition to the above, there are many studies related to emotional expression on social media, such as emotion causes detection [26], gender differences [27], and personality differences [28].

In general, with the rapid growth of social media users, expression of emotion has increased massively on these platforms. Social media now is a valuable resource for researchers studying emotions. Additionally, the significant role of emotional expression in social media has attracted more and more academic attention. Sentiment analysis based

on social media data, in particular, has become a hot topic in recent years.

B. CONSTRUCTION OF AN EMOTION LEXICON

As mentioned before, the lexicon-based approach is one of the main approaches for sentiment analysis [29], which calculates orientation for a document from the semantic orientation of words or phrases in the document [30]. And emotion lexicons are vital to sentiment analysis since they serve as the evidential basis of emotion [31]. Early emotion lexicons were created manually [12], [32], [33]. For instance, Stone *et al.* used existing dictionaries (Lasswell's Dictionary and Harvard IV-4 Dictionary) to create their General Inquirer Lexicon [32]. The emotion words were collected manually, and the relevant emotion polarities were annotated by hand. This method is labor-intensive, however, and it is not favored by researchers [31]. With the development of computer technology, semi-automatic methods [34], [35] and automatic methods [36] were widely used in the construction of emotion lexicons.

Producing a dictionary automatically via association (dictionary-based approach) was a main approach for lexicon construction [31]. This method first needs to collect a list of emotion words (also called seed words) and their known emotion polarities or categories. After searching for synonyms or antonyms of the seed word, calculating the association between the new word and the seed word, and formulating a judgment standard, the new words that meets the standard would be automatically or semi-automatically added to the lexicon. In this way, more new words could be discovered, and the emotion polarity or category they represent can also be inferred. Many emotion or sentiment lexicons were constructed using this method, such as SentiWordNet [35], [37] and HowNet [13]. However, since seed words and new words are usually taken from existing dictionaries or databases, the word list is not entirely suitable for all contexts [38]. In addition, lexicon publishing time also affects the use of the dictionary because of the lack of recent words and expressions. To improve the lexicon's effect of specific context, combining emotion word lists with scene data is advocated. For instance, Xu *et al.* utilized a large unlabeled corpus from Sogou, a Chinese search engine, to correct and expand two popular emotion lexicons (HowNet and NTUSD) [39]; Liu, Lei, and Wang constructed a domain-specific emotion lexicon by combining a context-sensitive emotion lexicon and the existing emotion word list [40]; and Wu *et al.* constructed an emotion lexicon termed "iSentiDictionary" by applying a self-training spreading method to expand emotion words from ConceptNet [41].

One drawback of existing emotion lexicons is that they do not consider the complexity of emotion categories, and most of them are only classified based on the polarity [31]. Even if the discrete emotions were considered, the selection of emotion categories was subjective without theoretical support [42]–[44]. This means that the lexicon built on their emotion categories is not comprehensive. In addition, previous

studies have pointed out that as time and contexts change, the applicability of a lexicon will also deteriorate [31], [38]. Because early emotion lexicons may lack current words and expressions, people's emotion expressions may also differ on different social networks [15]. Considering the above issues, we advocated an emotion lexicon construction method as follows: (a) using grounded theory approach to build a theoretical model based on the target context (Weibo in this study); (b) using the dimension of the model as the basis for the emotion category and extracting seed words based on the category; and (c) using machine learning algorithms to expand the emotion lexicon and manually filter the expanded results. This method is an extension of our previous work of corpus construction [45].

C. LEXICON-BASED APPROACH FOR SENTIMENT ANALYSIS

The lexicon-based approach is an unsupervised approach for sentiment analysis. First, it consists of emotion polarity or category for each word, to label the emotions of words in a document. Second, the emotion of the document is computed by aggregating the emotion of each word based on different approaches, like direct summation, weight calculation, and syntactic rules.

Many researchers have used a lexicon-based approach for sentiment analysis. Zhang *et al.* proposed a rule-based approach with two steps: 1) judging emotion of each sentence first based on word dependency; and 2) synthesizing the sentences' emotions to generate the emotion of the document [46]. Similarly, Taboada *et al.* calculated the emotion scores of words, phrases, sentences, and passages by specific rules and judged the documents' emotions based on different thresholds [29]. Rao *et al.* built an emotional dictionary for emotional tracking of online news, advocating an efficient algorithm and three pruning strategies [47]. Different from typical lexicon-based approaches, Saif *et al.* presented SentiCircles for sentiment analysis, which considered the co-occurrence patterns of words in different contexts in tweets to capture their semantics and update their pre-assigned strength and polarity in emotion lexicons accordingly [48].

Lexicon-based approaches for sentiment analysis on Weibo (also called Chinese Micro-Blog in some research) have continued to improve in recent years. For instance, Zhang *et al.* presented a sentiment analysis method for Weibo text based on the emotion lexicon to better support the work of network regulators [49]. The work of [4] constructed an extended emotion lexicon that contains the basic emotion words, the field emotion words, and the polysemic emotion words, all of which improve the accuracy of Weibo sentiment analysis. Wu *et al.* proposed a method for constructing multiple sentiment dictionaries, including the original sentiment dictionary, emoji dictionary, and other related dictionaries [50]. They then analyzed semantic rule sets between Weibo texts, took the inter-sentence analysis rules, and sentenced pattern analysis rules into the sentiment analysis of Weibo,

which further improves the accuracy of sentiment analysis. In addition, emotion lexicons have also been applied to the sentiment analysis of different data sources, such as movie reviews [51], [52], students' feedback [53], book reviews [54], and party statements [55]. In this study, we applied a rule-based approach for Weibo sentiment analysis and compared our emotion lexicon with existing lexicons based on the results of emotion recognition.

III. STUDY1: USING GROUNDED THEORY TO BUILD A WEIBO-BASED EMOTIONAL EXPRESSION MODEL

In study 1, we used the grounded theory method [16] and the social constructivist approach of [56] to build a model of Weibo emotional expression. Taking a bottom-up construction of a model or theory, this method is based on or grounded in the reality of people's experience. Drawing on the procedures of [57], our modeling process included data collection, data coding, and theoretical saturation test.

A. DATA COLLECTION

Since the number of words per Weibo text was limited to 140, we screened and collected a total of 7,535 Weibo texts expressing user emotions for subsequent coding. Qualitative research needs to select samples that can provide as much information as possible for research purposes [58], so we first used homogenous intensity sampling to deliberately choose Weibo texts which expressed various emotions. As a result, 1,000 Weibo texts from 23 active Weibo users were selected. After the process of open coding of the above texts, we used theoretical sampling [59] by targeting 141 new users who frequently expressed their emotions on Weibo. The sampling of additional data was directed by the evolving theoretical constructs, and a total of 5,633 new Weibo texts from the above users were collected in this process. Finally, 902 additional texts were collected for the theoretical saturation test. To avoid bias caused by user-specific expressions, the maximum number of Weibo texts per user was 50. In addition, since the users' personal information is not public, we selected different types of users based only on the topics they care about.

B. DATA CODING AND ANALYSIS

In this part, substantive coding, including open, axial, and selective coding [45], was first conducted to build the preliminary theoretical model of users' emotional expression on Weibo, and a total of 6,633 texts (88.03%) were coded in this step. Then, the remaining 902 texts were used for the theoretical saturation test. Additionally, consensual qualitative research [60], [61] was chosen for this study to analyze the text, reducing the subjectivity of individual coders. Texts were coded by a psychology graduate student and a PhD candidate, and each step was discussed in order to reach a consensus.

1) SUBSTANTIVE CODING

We first conducted open coding by examining the texts, sentence by sentence, to form initial concepts and cate-

TABLE 1. The coding categories for users' emotional expression on Weibo and frequencies for selective coding categories.

Selective coding (Chinese)	Axial coding (Chinese; Open coding numbers)	Sub-total for selective codes
Joy (快乐)	Happiness (幸福; 19), enjoyment (享受; 12), excitement (激动; 9), moved (感动; 10), contentment (满足; 11), pride (自豪; 4), boast (自夸; 3).	1497 (19.99%)
Expectation (期待)	Blessing (祝福; 15), encouragement (鼓励; 33), prayer (祈祷; 11), hope (希望; 19).	884 (11.81%)
Love (喜爱)	Praise (赞美; 38), recognition (认可; 15), gratitude (感谢; 20), respect (尊敬; 19), congratulation (恭喜; 27).	1296 (17.31%)
Anger (愤怒)	Resentment (怨恨; 8), fury (暴怒; 6), anger (生气; 46), hostility (敌意; 9)	643 (8.59%)
Anxiety (焦虑)	Distress (烦恼; 11), suffer (煎熬; 18), fear (恐惧; 28), irritation (烦躁; 7), escape (逃避; 6)	735 (9.82%)
Disgust (厌恶)	Scorn (轻蔑; 15), disdain (鄙视; 20), hate (讨厌; 21), loathing (憎恶; 9), antipathy (反感; 11)	841 (11.23%)
Sadness (悲伤)	Disappointment (失望; 14), loneliness (孤独; 7), regret (后悔; 5), boredom (无聊; 9), awkwardness (尴尬; 6), self-pity (自怜; 4), guilty (愧疚; 8), grievance (委屈; 11), helpless (无奈; 5),	1218 (16.27%)
Surprise (惊讶) 48	Thrill (惊险; 5), shock (震惊; 7), wonder (惊奇; 8), exclaim (惊叹; 8), pleasant surprise (惊喜; 3), unexpected (出乎意料; 12), doubt (疑惑; 5)	373 (4.98%)
Total		7487 (100%)

gories of emotion expression, such as source of happiness (Chinese: 快乐源泉), restless (Chinese: 心神不宁), getting better and better (Chinese: 越来越好), and caught off guard (Chinese: 猝不及防). Five hundred ninety-seven nodes were totally produced and examined for their inter-connections and further coded for 46 main categories (axial coding). For instance, happiness (Chinese: 幸福) is one of the main categories, including ha-ha (Chinese: 哈哈), so happy (Chinese: 好开心), and happy (Chinese: 开森; a special expression of happy in Chinese social networks). Finally, we performed selective coding to integrate the categories and formed the initial theoretical model, which included eight core categories of emotion expression on Weibo: joy, expectation, love, anger, anxiety, disgust, sadness, and surprise (see Table 1).

2) THEORETICAL SATURATION TEST

Theoretical saturation test is the last step in grounded theory methodology. In this step, additional data analysis will be performed continuously until theoretical saturation is reached (i.e., additional analysis no longer contributes to anything new about our theory). We examined the remaining 902 Weibo texts to perform the saturation test of our results, and the results show no new categories. This suggested that our theoretical model of emotional expression on Weibo was basically saturated [56]. Combining the coding results from the theoretical saturation test with our previous results, the final coding results are shown in Table 1.

C. RESULTS AND DISCUSSION

The coding results showed eight core categories of emotional expression on Weibo (see Table 1): joy, expectation, love,

anger, anxiety, disgust, sadness, and surprise. Our categories coincide with some of the basic emotions proposed in previous researches, such as the six basic emotions (happiness, sadness, anger, fear, disgust, and surprise) identified by Ekman in [62]; and the ten basic emotions (interest, joy, surprise, sadness, fear, shyness, guilt, angry, disgust, and contempt) presented by Izard in [63], [64]. This reminds us that emotions expressed on social media and basic emotions may not map onto one another in a 1-1 fashion.

In all emotion categories, we found that joy nearly accounted for a fifth of the texts, followed closely by love (17.31%) and sadness (16.27%). Among them, joy and sadness are generally considered as basic emotions [62]–[64]; so, it is interpretable that they occupy a vital position in emotional expressions on social media. Love was considered a complex emotion, intermingled with joy and trust [65]. Previous researchers have shown that love is an important topic in social networks [27], [66]. In addition to the above three emotions, expectation (11.81%) and disgust (11.23%) were the most expressed emotions. Similarly, disgust is a basic emotion, but expectation is relatively rarely mentioned in studies. In fact, Quan and Ren recruited eleven annotators to mark the emotions expressed in a large number of Weibo texts and found that expectation is one of the most common emotions, similar to our results [43].

Each emotion accounted for less than 10% of the texts. Although expressed less in this study, the expression of anxiety tends to erupt during disasters [67]. And previous research has also shown that anxiety can be detected through social platforms, such as Facebook [68]. Anger is also an

TABLE 2. Examples of text segmentation results.

Chinese	English translation
[我][发誓][我][再也不][熬夜][了]	[I][swear][I][won't][stay up all night][Chinese modal particle]
[为][毛][我][的][空调][是][制热][的][模式]	[Why][my][air conditioner][is][heating][mode]
[这][便是][执子之手][与子偕老][吧]	[This][is][holding your hand][to grow old with you][Chinese modal particle]

emotion that will erupt in the short term. Fan *et al.* found that anger spread faster and more widely on Weibo than other emotions because it was shared by both close connections and distant connections [69], whereas an emotion like joy was shared mainly by close connections [70]. Finally, surprise was the least expressed emotion in this study (4.98%), but it was widely expressed in other social platforms, such as Twitter [71].

In summary, Study 1 used grounded theory to build a theoretical model of Weibo emotional expression. It showed a similar emotion category as the basic emotions, but also contained a few complex emotions. In Study 2, we used this model to construct an emotion lexicon for sentiment analysis, and the results of coding was referenced to form the seed words.

IV. STUDY 2: CONSTRUCTING AN EMOTION LEXICON BASED ON THE SEMI-AUTOMATIC METHOD

In Study 2, we constructed an emotion lexicon based on the core categories of emotional expression on Weibo. Using a semi-automatic method, five steps were conducted as follows: constructing a seed word set, collecting Weibo text data, expanding words, filtering the expanded lexicon manually, and data preprocessing.

A. CONSTRUCTION OF SEED WORD SET

In this part, our goal was to build a seed word set for each emotion category. Referring to the method of [45], we used 597 nodes, 46 main categories, and 8 core categories, all of which were identified by using grounded theory in Study 1 as original words. Then, we manually checked these words to remove ambiguous words and antonyms. As a result, a total of 496 words were retained as seed words.

B. DATA COLLECTION

We collected a large number of texts in Weibo for word expansion. The total data included about 1.01 million Weibo texts since the expansion result was basically stable when the training text exceeded seven hundred thousand.

C. DATA PREPROCESSING

Two steps were conducted in this part. First, we filtered unwanted information elements (e.g., emoticons, web links, mention of someone's symbol) in Weibo text based on some rules. Second, text segmentation was performed to cut a sequence of Chinese characters into individual words.

1) FILTERING OF UNWANTED INFORMATION ELEMENTS

Weibo text often appears with some emoticons, pictures, Web links, someone's @ symbol, and other information elements, which bring richness and color to Weibo texts; however, this colorful text also presents some difficulties for some researchers [49]. Therefore, it is necessary to filter the Weibo text to facilitate the research work. We aimed to build a new emotion lexicon, so five aspects were considered as follows:

- 1) Filtering "@ + username." The symbol "@ + username" in Weibo is used to tell someone something or to get someone's attention, and it has no substantive effect on emotional expression.
- 2) Filtering Web links, animation, videos, pictures, and emoticons. Although these information elements may be related to emotional expression, this study focused only on textual information. Therefore, they were filtered.
- 3) Filtering location information. Weibo has the function of sharing positioning information, but this is useless for sentiment analysis.
- 4) Removing punctuation because punctuation may affect the performance of subsequent models.
- 5) Filtering blank lines. After the above operations, some Weibo information was completely removed, thus forming blank lines. All the blank lines were then removed.

2) TEXT SEGMENTATION

A total of 1,012,357 Weibo texts were retained after filtering, and these texts were then split into individual words for subsequent modeling. The Jieba Chinese word segmentation tool [72] was applied for text segmentation. This tool has been widely used in previous research [73], [74]. Table 2 presents three examples of text segmentation results.

D. WORDS EXPANSION

In this part, we aimed to search for words similar to the seed word by calculating the association between them. Therefore, we first applied Word2Vec to construct a word vector model. Word2Vec is an open source tool, which can train a set of word vectors from massive corpus (1.01 million Weibo texts in this study) through a neural network model [75]. The core function parameters set in this study were as follows: $sg = 0$, $size = 200$, $window = 5$, and $min_count = 1e-3$. After model training, each word could be represented as a vector. Then, a cosine similarity measure was conducted to calculate the association between the seed word and the new word.

TABLE 3. Partial subset word list of seed word “ha-ha-ha”.

Chinese	English translation	Cosine	Chinese	English translation	Cosine
哈哈哈哈哈	Ha-ha-ha-ha	0.923	太逗	Too funny	0.546
哈哈	Ha-ha	0.864	笑死	Laugh my ass off	0.524
啊哈哈哈哈哈	A-ha-ha-ha-ha	0.765	痴汉笑	Silly laugh	0.507
哈哈笑	Laugh with haha	0.731	三声	Laugh three times	0.504
噗	Titter	0.631	好好笑	Very funny	0.482

TABLE 4. Final number of words or phrases contained in each emotion category.

Joy	Expectation	Love	Anger	Anxiety	Disgust	Sadness	Surprise
428	357	621	396	382	379	461	130

The cosine between each seed word vector and each new word vector was measured, and the 400 words most similar to each seed word were retained as a subset of this word. For example, “ha-ha-ha (Chinese: 哈哈哈哈哈)” is one of the seed words, and part of its subset is shown in Table 3. Finally, the seed words and their subsets for each emotion category were merged, and repeated words in each category were removed.

E. MANUAL WORD FILTERING

In the last step, we manually checked the automatically expanded words because words or phrases not related to emotional expression were also included. For instance, “laugh three times” is similar to “ha-ha-ha” (see Table 3); but “three times” is not meaningful for expressing joy. “Laugh” was already included in the word list, so the phrase “laugh three laughs” was removed. This step was conducted by three psychology graduate students, and each word or phrase was discussed in order to reach a consensus. The final number of words or phrases in each emotion category is shown in Table 4.

F. RESULTS AND DISCUSSION

In Study 2, we used a semi-automatic method to construct a Weibo-based emotion lexicon. In order to present the words contained in our lexicon intuitively, we counted the frequency of all the words appearing in the 1.01 million Weibo texts; and the 100 most frequently occurring words in each emotion category were then shown in word clouds (see Fig.1). For instance, “hope” (Chinese: 希望), “wish” (Chinese: 愿), and “come on” (Chinese: 加油) were the most common words or phrases that express the emotion of expectation. They appeared 21,850, 20,587 and 19,912 times in the 1.01 million Weibo texts, respectively.

Although our lexicon contained fewer words (a total of 2,964 words), a lexicon constructed from a specific scene corpus is often more effective for sentiment analysis of the current context [38]. In addition, our lexicon focused on emotions expressed by people, so we believe it can play a useful role in the recognition of emotional expressions on Weibo. Therefore, we also tested the effectiveness of this lexicon in sentiment analysis in Study 3.

V. STUDY3: RECOGNIZING THE EMOTIONS EXPRESSED IN WEIBO TEXT

In Study 3, we applied a lexicon-based approach for Weibo sentiment analysis. We used different lexicons and the same method for sentiment analysis and compared the effects of different emotion lexicons on Weibo emotion recognition from the perspectives of coverage and accuracy.

A. EMOTION LEXICONS FOR COMPARISON

We selected five popular Chinese emotion or sentiment lexicons for comparison as follows: SC-LIWC [11], NTUSD [12], HowNet [13], THUOCL [14], and Affective Lexicon Ontology (ALO) [76]. As shown in Table 5, some lexicons only contain emotion polarity, so the subsequent analysis was conducted to recognize both emotion category and emotion polarity of Weibo text. In the analysis of emotion polarity, we merged our lexicon as follows: joy and love were merged into positive emotions and anger, anxiety, disgust, and sadness were merged into negative emotions; surprise and expectation were not used in the analysis of emotion polarity.

B. DATA COLLECTION AND EMOTION ANNOTATION

We collected a total of 7,515 Weibo texts, other than those used in Study 1 and Study 2, as targets for emotion recognition. These texts were then annotated manually with emotion category and polarity by three psychology graduate students, and each text was discussed to reach a consensus. These annotated texts were used as ground truth data for subsequent tests.

C. EMOTION RECOGNITION

After data preprocessing (same as Study2), a rule-based approach was applied to recognize the emotion category or polarity. With reference to the methods of previous research [77], [78], recognition of emotion expressed in Weibo text was conducted in the following way:

- 1) Each sentence of Weibo text is examined for the presence of words or phrases belonging to different emotion categories or polarities.



FIGURE 1. Word cloud of top 100 emotion words in each emotion category.

TABLE 5. Emotion categories and word numbers in different lexicons.

Lexicon	Positive	Negative	Others	Total
SC-LIWC	564	Anxiety (129), anger (282), sadness (165), others (248)	None	1612
NTUSD	2647	7741	None	10388
HowNet	4566	4370	None	8936
THUOCL	5567	4468	None	10035
ALO	Joy (1966), love (1013)	Anger (388), sadness (2315), disgust (10282), anxiety (1179)	Surprise (228)	17173
Our lexicon	Joy (428), love (621)	Anger (396), anxiety (382), disgust (379), sadness (461)	Surprise (130), expectation (357)	2964

- 2) If the sentence is found, it will be checked for the presence of negation words, which may influence the judgment of emotional expression. A total of thirty Chinese negation words, summarized in SC-LIWC [11], is utilized in this step.
- 3) If no negation word is found, the emotion of the entire found sentence will be detected according to the emotion word.
- 4) If a negation word is found, the negation word and emotion words will be checked for their position in the sentence.
- 5) If the negation word appears before the emotion word, and the distance between them is less than 3 words, the corresponding emotion word will be invalid.
- 6) If the negation word appears before the emotion word, but the distance is more than 2 words, the emotion of the

TABLE 6. Examples of emotion recognition results.

Original Weibo text	English translations	Emotion categories and polarity
好[开心]好[开心]我家狗狗找到了	So [happy] So [happy] My dog is found	Joy, positive
[妈的]咋就这么[憋屈]呢	[Oh shit] I'm so [aggrieved]	Anger, sadness, negative
你[居然]要看体育新闻	You [actually] want to watch sports news	Surprise

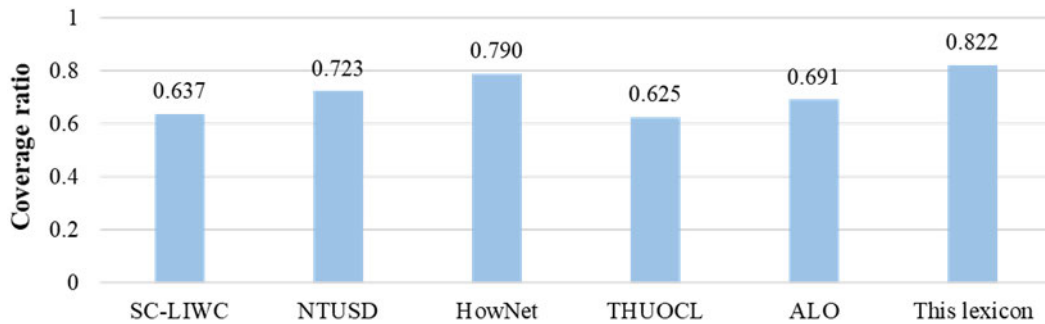


FIGURE 2. Coverage ratios of different Chinese emotion lexicons.

TABLE 7. Emotion polarity recognition results of different lexicons.

Lexicon	Positive			Negative		
	Precision	Recall	F1	Precision	Recall	F1
SC-LIWC	0.669	0.581	0.622	0.894	0.471	0.617
NTUSD	0.679	0.595	0.634	0.839	0.637	0.724
HowNet	0.474	0.735	0.576	0.781	0.540	0.638
THUOCL	0.637	0.604	0.620	0.878	0.439	0.585
ALO	0.640	0.647	0.643	0.786	0.561	0.655
This lexicon	0.841	0.773	0.806	0.943	0.691	0.798

sentence will be recognized according to the emotion word.

- 7) If the negation word appears after the emotion word, the emotion of the sentence will be recognized according to the emotion word.

Based on the above rules, most of Weibo texts were given labels for multiple emotion categories and polarity, but not multiple times for the same labels. And three samples of the output are shown in Table 6. Each emotion lexicon performed the above operations separately, and the comparison result of Weibo text emotion recognition was finally presented in section D.

D. RESULTS AND DISCUSSION

We compared the effectiveness of each lexicon from the perspectives of coverage and accuracy.

1) COVERAGE RESULTS

Regardless of the accuracy of emotion recognition, we calculated the coverage ratio of each lexicons as follows: $Coverage\ ratio = n/N * 100\%$, where n is the number of Weibo texts that match the lexicon's emotion words, and N is the total number of Weibo texts. As shown in Fig. 2,

the coverage of our lexicon is 4.05% to 31.52% ahead of other lexicons.

2) EMOTION RECOGNITION RESULTS

Although we used an unsupervised approach for emotion recognition, emotions expressed in the 7,515 Weibo texts had been manually annotated (ground truth). Therefore, we used precision, recall, and F1 measure as the criteria to evaluate the result of emotion recognition. The results of recognition of emotion polarity and emotion category are presented in Table 7 and Table 8, respectively. For the recognition of emotion polarity, results based on our lexicon are better than other lexicons. And our lexicon is basically better in the recognition of emotion categories, except for anxiety.

3) DISCUSSION

We believe that different training corpus and differences in emotion definitions are the two core reasons for the above results. First, ALO, HowNet, and THUOCL applied written corpus, including books, prose, and Chinese dictionary, to expand their emotion words. This is somewhat different from corpus based on social networks because online

TABLE 8. Emotion category recognition results of different lexicons.

Emotion category	This lexicon			ALO			SC-LIWC		
	Precision	Recall	F1	Precision	Recall	F1	Precision	Recall	F1
Joy	0.667	0.716	0.692	0.599	0.485	0.536	-	-	-
Expectation	0.468	0.750	0.576	-	-	-	-	-	-
Love	0.543	0.704	0.613	0.289	0.591	0.388	-	-	-
Anger	0.582	0.517	0.548	0.469	0.051	0.091	0.311	0.265	0.286
Anxiety	0.533	0.644	0.583	0.589	0.261	0.361	0.600	0.377	0.463
Disgust	0.602	0.479	0.534	0.230	0.551	0.324	-	-	-
Sadness	0.710	0.535	0.610	0.671	0.324	0.437	0.677	0.233	0.346
Surprise	0.413	0.638	0.502	0.518	0.233	0.322	-	-	-

expressions are special. SC-LIWC has used a network corpus to expand the words, but the number is still small. Second, our lexicon focuses on emotions expressed by people, while the other lexicons confuse words used to describe emotions with words used to express emotions. Therefore, although the number of emotion words in this study is smaller, the coverage rate and accuracy are higher.

VI. GENERAL DISCUSSION AND CONCLUSION

By combining grounded theory and semi-automatic methods, we built a Weibo-based emotion lexicon for sentiment analysis. We first took a bottom-up approach to derive a theoretical model for emotions expressed on Weibo. The substantive coding led to eight core categories of discrete emotions, which were then used to construct our emotion lexicon. Second, we built our emotion lexicon by manually selecting seed words, constructing a word vector model to expand words, and creating rules to filter words. As a result, a new emotion lexicon containing 2,964 words was created for analysis of emotion expressed on Weibo. Finally, we applied a lexicon-based approach to recognize the emotions expressed in Weibo text. Compared to five other Chinese lexicons, our lexicon demonstrated better performance in Weibo emotion recognition.

Study 1 used grounded theory to construct a typology of emotions expressed on Weibo and to search seed words for the emotion lexicon, which provides an example of using qualitative research methods to facilitate sentiment analysis. As mentioned before, most of the existing emotion lexicons are only classified based on the polarity [31], and the other multi-category emotion lexicons only subjectively select several discrete emotions (e.g., SC-LIWC and ALO). While grounded theory approach is a useful qualitative research method for constructing theory, which is widely applied in the fields of psychology and social sciences. Use of grounded theory makes up for the lack of theoretical foundation, allowing us to see more fully the categories of emotions people express on Weibo. In fact, grounded theory is widely used in research about emotions [79], [80] and other areas, such as personality [81], need [82], and psychological resilience [83]. A combination of qualitative research and artificial

intelligence methods has also been advocated in recent years [45]. Therefore, in Study 2, we used a semi-automated method that combines the results of grounded theory, natural language processing (NLP), and manual screening to build the lexicon. Among them, the grounded theory provided us with emotion categories and corresponding seed words. Calculating word vector similarity, widely used in NLP [75], helped us automatically search similar words and expand the lexicon. And a manual inspection approach corrected errors of machine expansion. As a result, a novel emotion lexicon for analysis of emotions expressed on Weibo was created. The method used in this study can also be generalized to the construction of other lexicons by replacing the seed words and corpus according to the goal. In Study 3, we used a lexicon-based method for sentiment analysis to test the effectiveness of our lexicon. The results showed that our lexicon performs well in Weibo sentiment analysis and is superior to the other five common Chinese lexicons. The main reason for the above results is that we created an emotion lexicon specifically for the Weibo context. To consider the context of words can effectively improve the performance of sentiment analysis [4], [38] as well as avoid some cross-context problems, such as the fuzziness of the intensity-of-word sentiment polarity [84].

Our work indirectly improved the performance of sentiment analysis in Weibo text and provided a reference for other contexts. As one of the vital foundations of sentiment analysis, emotion lexicons serve as the evidential basis of emotion [31]. The emotion category and corresponding emotion words suitable for the current not only facilitate sentiment analysis but also support other studies derived from social media sentiment analysis, such as emotional contagion [23], [24], emotion cause detection [26], and personality differences [28].

This study has several notable limitations. First, all our data was taken from Chinese Weibo users. If our lexicon is used for sentiment analysis of other social media platforms, it is necessary to modify and expand the lexicon based on the corpus of the target platform. Second, our work is on-going and labor-intensive. If we want to update the lexicon in the future, such as by adding new network terms, it could take

considerable time and manpower. Third, our lexicon did not include the strength of the association between each word and the emotion category. Emotion lexicons containing words and their relevant emotion values may be better utilized for sentiment analysis, such as SentiWordNet [37], SenticNet 5 [85], and CSenticNet [86]. Finally, the sentiment analysis method used in this study is not the strongest, and it does not make full use of the lexicon. In fact, social text sentiment analysis technology is constantly being updated [4], [49], [50]; and continuous updating methods to achieve more accurate sentiment analysis is crucial in future research.

REFERENCES

- [1] D. Derks, A. H. Fischer, and A. E. R. Bos, "The role of emotion in computer-mediated communication: A review," *Comput. Hum. Behav.*, vol. 24, no. 3, pp. 766–785, May 2008.
- [2] X. Hu, L. Tang, J. Tang, and H. Liu, "Exploiting social relations for sentiment analysis in microblogging," in *Proc. 6th ACM Int. Conf. Web Search Data Mining (WSDM)*, 2013, pp. 537–546.
- [3] L. Huang, S. Li, and G. Zhou, "Emotion corpus construction on microblog text," in *Chinese Lexical Semantics*, vol. 9332, Q. Lu and H. H. Gao, Eds. 2015, pp. 204–212.
- [4] G. Xu, Z. Yu, H. Yao, F. Li, Y. Meng, and X. Wu, "Chinese text sentiment analysis based on extended sentiment dictionary," *IEEE Access*, vol. 7, pp. 43749–43762, 2019.
- [5] Y. Wang, S. Feng, D. Wang, G. Yu, and Y. Zhang, "Multi-label Chinese microblog emotion classification via convolutional neural network," in *Proc. Asia-Pacific Web Conf.*, vol. 9931, 2016, pp. 567–580.
- [6] J. C. Eichstaedt, R. J. Smith, R. M. Merchant, L. H. Ungar, P. Crutchley, D. Preotiuc-Pietro, D. A. Asch, and H. A. Schwartz, "Facebook language predicts depression in medical records," *Proc. Nat. Acad. Sci. USA*, vol. 115, no. 44, pp. 11203–11208, Oct. 2018.
- [7] B. Liu and L. Zhang, "A survey of opinion mining and sentiment analysis," in *Mining Text Data*. Cham, Switzerland: Springer, 2012, pp. 415–463.
- [8] J. Bollen, H. Mao, and X. Zeng, "Twitter mood predicts the stock market," *J. Comput. Sci.*, vol. 2, no. 1, pp. 1–8, Mar. 2011.
- [9] S. M. Liu and J.-H. Chen, "A multi-label classification based approach for sentiment classification," *Expert Syst. Appl.*, vol. 42, no. 3, pp. 1083–1093, Feb. 2015.
- [10] B. Liu, "Sentiment analysis and opinion mining," *Synth. Lectures Hum. Lang. Technol.*, vol. 5, no. 1, pp. 1–167, 2012.
- [11] C.-L. Huang, C. K. Chung, N. Hui, Y.-C. Lin, Y.-T. Seih, B. C. P. Lam, W.-C. Chen, M. H. Bond, and J. W. Pennebaker, "The development of the Chinese linguistic inquiry and word count dictionary," *Chin. J. Psychol.*, vol. 54, no. 2, pp. 185–201, 2012.
- [12] L.-W. Ku, Y.-T. Liang, and H.-H. Chen, "Opinion extraction, summarization and tracking in news and blog corpora," in *Proc. AAAI Spring Symp. Comput. Approaches Analyzing Weblogs*, 2006, pp. 100–107.
- [13] Z. Dong, Q. Dong, and C. Hao, "Hownet and its computation of meaning," in *Proc. 23rd Int. Conf. Comput. Linguistics Demonstrations*, 2010, pp. 53–56.
- [14] S. Han. (2016). *THUOCL: Tsinghua Open Chinese Lexicon*. Accessed: Jun. 6, 2020. [Online]. Available: <http://nlp.csai.tsinghua.edu.cn/site2/index.php/13-sms>
- [15] S. F. Waterloo, S. E. Baumgartner, J. Peter, and P. M. Valkenburg, "Norms of online expressions of emotion: Comparing Facebook, Twitter, Instagram, and WhatsApp," *New Media Soc.*, vol. 20, no. 5, pp. 1813–1831, May 2018.
- [16] B. G. Glaser, A. L. Strauss, and E. Strutzel, "The discovery of grounded theory; strategies for qualitative research," *Nursing Res.*, vol. 17, no. 4, p. 364, Jul. 1968.
- [17] D. I. Tamir and J. P. Mitchell, "Disclosing information about the self is intrinsically rewarding," *Proc. Nat. Acad. Sci. USA*, vol. 109, no. 21, pp. 8038–8043, May 2012.
- [18] S. Sprecher, S. Treger, J. D. Wondra, N. Hilaire, and K. Wallpe, "Taking turns: Reciprocal self-disclosure promotes liking in initial interactions," *J. Exp. Social Psychol.*, vol. 49, no. 5, pp. 860–866, Sep. 2013.
- [19] A. Carpenter and K. Greene, "Social penetration theory," *Int. Encyclopedia Interpersonal Commun.*, pp. 1–4, Dec. 2015, doi: [10.1002/9781118540190.wbeic160](https://doi.org/10.1002/9781118540190.wbeic160).
- [20] K. G. Meleshko and L. E. Alden, "Anxiety and self-disclosure: Toward a motivational model," *J. Pers. Soc. Psychol.*, vol. 64, no. 6, pp. 1000–1009, Jun. 1993.
- [21] J. Frattaroli, "Experimental disclosure and its moderators: A meta-analysis," *Psychol. Bull.*, vol. 132, no. 6, p. 823, 2006.
- [22] J. B. Walther, "Computer-mediated communication: Impersonal, inter-personal, and hyperpersonal interaction," *Commun. Res.*, vol. 23, no. 1, pp. 3–43, Feb. 1996.
- [23] A. D. I. Kramer, J. E. Guillory, and J. T. Hancock, "Experimental evidence of massive-scale emotional contagion through social networks," *Proc. Nat. Acad. Sci. USA*, vol. 111, no. 24, pp. 8788–8790, Jun. 2014.
- [24] E. Ferrara and Z. Yang, "Measuring emotional contagion in social media," *PLoS ONE*, vol. 10, no. 11, Nov. 2015, Art. no. e0142390.
- [25] W. J. Brady, J. A. Wills, J. T. Jost, J. A. Tucker, and J. J. Van Bavel, "Emotion shapes the diffusion of moralized content in social networks," *Proc. Nat. Acad. Sci. USA*, vol. 114, no. 28, pp. 7313–7318, Jul. 2017.
- [26] K. Gao, H. Xu, and J. Wang, "A rule-based approach to emotion cause detection for Chinese micro-blogs," *Expert Syst. Appl.*, vol. 42, no. 9, pp. 4517–4528, Jun. 2015.
- [27] M. Thelwall, D. Wilkinson, and S. Uppal, "Data mining emotion in social network communication: Gender differences in MySpace," *J. Amer. Soc. Inf. Sci. Technol.*, vol. 61, no. 1, pp. 190–199, Jan. 2010.
- [28] D. Dupré, N. Andelic, A. Zajac, G. Morrison, and G. McKeown, "The effect of personality and social context on willingness to share emotion information on social media," *PsyArXiv*, to be published. [Online]. Available: <https://psyarxiv.com/g8ndz/>, doi: [10.31234/osf.io/g8ndz](https://doi.org/10.31234/osf.io/g8ndz).
- [29] M. Taboada, J. Brooke, M. Tofiloski, K. Voll, and M. Stede, "Lexicon-based methods for sentiment analysis," *Comput. Linguistics*, vol. 37, no. 2, pp. 267–307, Jun. 2011.
- [30] P. D. Turney, "Thumbs up or thumbs down?: Semantic orientation applied to unsupervised classification of reviews," in *Proc. 40th Annu. Meeting Assoc. Comput. Linguistics (ACL)*, 2001, pp. 417–424.
- [31] H. Peng, E. Cambria, and A. Hussain, "A review of sentiment analysis research in chinese language," *Cognit. Comput.*, vol. 9, no. 4, pp. 423–435, Aug. 2017.
- [32] P. J. Stone, D. C. Dunphy, and M. S. Smith, *The General Inquirer: A Computer Approach to Content Analysis*. Cambridge, MA, USA: MIT Press, 1966.
- [33] R. M. Tong, "An operational system for detecting and tracking opinions in on-line discussion," in *Proc. Work. Notes ACM SIGIR Workshop Oper. Text Classification*, vol. 1, 2001, pp. 1–6.
- [34] M. Hu and B. Liu, "Mining and summarizing customer reviews," in *Proc. ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2004, pp. 168–177.
- [35] A. Esuli and F. Sebastiani, "SENTIWORDNET: A publicly available lexical resource for opinion mining," in *Proc. LREC*, vol. 6, 2006, pp. 417–422.
- [36] G. Xu, X. Meng, and H. Wang, "Build Chinese emotion lexicons using a graph-based algorithm and multiple resources," in *Proc. 23rd Int. COLING*, 2010, pp. 1209–1217.
- [37] S. Baccianella, A. Esuli, and F. Sebastiani, "SENTIWORDNET 3.0: An enhanced lexical resource for sentiment analysis and opinion mining," in *Proc. LREC*, vol. 10, 2010, pp. 2200–2204.
- [38] Y. Su and S. Li, "Constructing Chinese sentiment lexicon using bilingual information," in *Proc. Workshop Chin. Lexical Semantics*, 2012, pp. 322–331.
- [39] H. Xu, K. Zhao, L. Qiu, and C. Hu, "Expanding Chinese sentiment dictionaries from large scale unlabeled corpus," in *Proc. 24th Pacific Asia Conf. Lang., Inf. Comput.*, 2010, pp. 301–310.
- [40] L. Liu, M. Lei, and H. Wang, "Combining domain-specific sentiment lexicon with hownet for Chinese sentiment analysis," *J. Comput.*, vol. 8, no. 4, pp. 878–883, Apr. 2013.
- [41] H.-H. Wu, A. C.-R. Tsai, R. T.-H. Tsai, and J. Y.-J. Hsu, "Building a graded Chinese sentiment dictionary based on commonsense knowledge for sentiment analysis of song lyrics," *J. Inf. Sci. Eng.*, vol. 29, no. 4, pp. 647–662, Jul. 2013.
- [42] S. Aman and S. Szpakowicz, "Identifying expressions of emotion in text," in *Proc. Text, Speech Dialogue*, vol. 4629, V. Matousek and P. Mautner, Eds. 2007, pp. 196–205.
- [43] C. Quan and F. Ren, "Construction of a blog emotion corpus for Chinese emotional expression analysis," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, vol. 3, 2009, pp. 1446–1454.
- [44] B. Ma, Y. Huang, and B. Deng, "Generating sentiment lexicon of online social network based on emotions," *Comput. Eng. Des.*, vol. 37, no. 5, p. 3, 2016.

- [45] Z. Sun, Z. Ji, P. Zhang, C. Chen, X. Qian, X. Du, and Q. Wan, "Automatic labeling of mobile apps by the type of psychological needs they satisfy," *Telematics Informat.*, vol. 34, no. 5, pp. 767–778, Aug. 2017.
- [46] C. Zhang, D. Zeng, J. Li, F.-Y. Wang, and W. Zuo, "Sentiment analysis of Chinese documents: From sentence to document level," *J. Amer. Soc. Inf. Sci. Technol.*, vol. 60, no. 12, pp. 2474–2487, Dec. 2009.
- [47] Y. Rao, J. Lei, L. Wenyin, Q. Li, and M. Chen, "Building emotional dictionary for sentiment analysis of online news," *World Wide Web*, vol. 17, no. 4, pp. 723–742, Jul. 2014.
- [48] H. Saif, Y. He, M. Fernandez, and H. Alani, "Contextual semantics for sentiment analysis of Twitter," *Inf. Process. Manage.*, vol. 52, no. 1, pp. 5–19, Jan. 2016.
- [49] S. Zhang, Z. Wei, Y. Wang, and T. Liao, "Sentiment analysis of Chinese micro-blog text based on extended sentiment dictionary," *Future Gener. Comput. Syst.*, vol. 81, pp. 395–403, Apr. 2018.
- [50] J. Wu, K. Lu, S. Su, and S. Wang, "Chinese micro-blog sentiment analysis based on multiple sentiment dictionaries and semantic rule sets," *IEEE Access*, vol. 7, pp. 183924–183939, 2019.
- [51] R. Kaur, "Sentiment analysis of movie reviews: A study of machine learning algorithms with various feature selection methods," *Int. J. Comput. Sci. Eng.*, vol. 5, no. 9, pp. 113–121, Sep. 2017.
- [52] K. Lu and J. Wu, "Sentiment analysis of film review texts based on sentiment dictionary and SVM," in *Proc. 3rd Int. Conf. Innov. Artif. Intell.*, 2019, pp. 73–77.
- [53] Q. Rajput, S. Haider, and S. Ghani, "Lexicon-based sentiment analysis of teachers' evaluation," *Appl. Comput. Intell. Soft Comput.*, vol. 2016, pp. 1–12, Oct. 2016, doi: 10.1155/2016/2385429.
- [54] C. Zhang, T. Tong, and Y. Bu, "Examining differences among book reviews from various online platforms," *Online Inf. Rev.*, vol. 43, no. 7, pp. 1169–1187, Nov. 2019.
- [55] M. Haselmayer and M. Jenny, "Sentiment analysis of political communication: Combining a dictionary approach with crowdcoding," *Qual. Quantity*, vol. 51, no. 6, pp. 2623–2646, Nov. 2017.
- [56] K. Charmaz, "A constructivist grounded theory analysis of losing and regaining a valued self," in *Five Ways of Doing Qualitative Analysis: Phenomenological Psychology, Grounded Theory, Discourse Analysis, Narrative Research, and Intuitive Inquiry*. New York, NY, USA: The Guilford Press, 2011, pp. 165–204.
- [57] J. M. Corbin and A. Strauss, "Grounded theory research: Procedures, canons, and evaluative criteria," *Qualitative Sociol.*, vol. 13, no. 1, pp. 3–21, 1990.
- [58] M. Q. Patton, *Qualitative Research and Evaluation Methods*. Thousand Oaks, CA, USA: Sage, 2002.
- [59] M. Stanley, "A grounded theory of the wellbeing of older people," in *Qualitative Research for Allied Health Professionals: Challenging Choices*. West Sussex, U.K.: Wiley, 2006, pp. 63–78.
- [60] C. E. Hill, B. J. Thompson, and E. N. Williams, "A guide to conducting consensual qualitative research," *Counseling Psychologist*, vol. 25, no. 4, pp. 517–572, Oct. 1997.
- [61] C. E. Hill, S. Knox, B. J. Thompson, E. N. Williams, S. A. Hess, and N. Ladany, "Consensual qualitative research: An update," *J. Couns. Psychol.*, vol. 52, no. 2, pp. 196–205, Apr. 2005.
- [62] P. Ekman, "An argument for basic emotions," *Cognition Emotion*, vol. 6, nos. 3–4, pp. 169–200, May 1992.
- [63] C. E. Izard, "Basic emotions, natural kinds, emotion schemas, and a new paradigm," *Perspect. Psychol. Sci.*, vol. 2, no. 3, pp. 260–280, Sep. 2007.
- [64] C. E. Izard, "Emotion theory and research: Highlights, unanswered questions, and emerging issues," *Annu. Rev. Psychol.*, vol. 60, no. 1, pp. 1–25, Jan. 2009.
- [65] R. Plutchik, "The nature of emotions: Human emotions have deep evolutionary roots, a fact that may explain their complexity and provide tools for clinical practice," *Amer. Scientist*, vol. 89, no. 4, pp. 344–350, 2012.
- [66] M. Yassine and H. Hajj, "A framework for emotion mining from text in online social networks," in *Proc. IEEE Int. Conf. Data Mining Workshops*, Dec. 2010, pp. 1136–1142.
- [67] S.-J. Baek, H. Jeong, and K. Kobayashi, "Disaster anxiety measurement and corpus-based content analysis of crisis communication," in *Proc. IEEE Int. Conf. Syst., Man, Cybern.*, Oct. 2013, pp. 1789–1794.
- [68] K. C. Fernandez, C. A. Levinson, and T. L. Rodebaugh, "Profiling: Predicting social anxiety from facebook profiles," *Social Psychol. Personality Sci.*, vol. 3, no. 6, pp. 706–713, Nov. 2012.
- [69] R. Fan, J. Zhao, Y. Chen, and K. Xu, "Anger is more influential than joy: Sentiment correlation in weibo," *PLoS ONE*, vol. 9, no. 10, Oct. 2014, Art. no. e110184.
- [70] R. C. Martin and L. E. Vieaux, "9. The digital rage: How anger is expressed online," in *The Psychology of Social Networking*, vol. 2. Warsaw, Poland: De Gruyter Open Poland, 2015, pp. 117–127, doi: 10.1515/9783110473858-011.
- [71] S. Kim, J. Bak, and A. H. Oh, "Do you feel what i feel? social aspects of emotions in twitter conversations," presented at the 6th Int. AAAI Conf. Weblogs Social Media, 2012.
- [72] J. Sun. (Jun. 2012). *Jieba Chinese Word Segmentation Tool*. [Online]. Available: <https://github.com/fxsjy/jieba>
- [73] L. Yan, B. Bai, W. Chen, and D. O. Wu, "New word extraction from chinese financial documents," *IEEE Signal Process. Lett.*, vol. 24, no. 6, pp. 770–773, Jun. 2017, doi: 10.1109/lsp.2017.2690599.
- [74] J. Du, X. Na, X. Liu, and H. Bu, "AISHELL-2: Transforming mandarin ASR research into industrial scale," 2018, *arXiv:1808.10583*. [Online]. Available: <http://arxiv.org/abs/1808.10583>
- [75] X.-Y. Song, L.-T. Jin, T. Liu, Y. Zhao, and Y. Sun, "Research on the construction of sentiment dictionary based on Word2vec," in *Proc. Int. Conf. Algorithms, Comput. Artif. Intell.*, Sanya, China, 2018, pp. 1–6.
- [76] L. Xu, H. Lin, Y. Pan, H. Ren, and J. Chen, "Constructing the affective lexicon ontology," *J. China Soc. Sci. Tech. Inf.*, vol. 27, no. 2, pp. 180–185, 2008.
- [77] D. Kan. (2012). *Rule-Based Approach to Sentiment Analysis at ROMIP 2011*. Accessed: Jun. 6, 2020. [Online]. Available: <http://www.dialog-21.ru/digests/dialog2012/materials/pdf/Kan.pdf>
- [78] A. Neviaroskaya, H. Prendinger, and M. Ishizuka, "Affect analysis model: Novel rule-based approach to affect sensing from text," *Natural Lang. Eng.*, vol. 17, no. 1, pp. 95–135, Jan. 2011.
- [79] D. T. Bishop, C. I. Karageorghis, and G. Loizou, "A grounded theory of young tennis players' use of music to manipulate emotional state," *J. Sport Exercise Psychol.*, vol. 29, no. 5, pp. 584–607, Oct. 2007.
- [80] E. A. Bolger, "Grounded theory analysis of emotional pain," *Psychotherapy Res.*, vol. 9, no. 3, pp. 342–362, Fall 1999.
- [81] L. Jia, Y. Xu, and M. Wu, "Campus job suppliers' preferred personality traits of Chinese graduates: A grounded theory investigation," *Social Behav. Personality Int. J.*, vol. 42, no. 5, pp. 769–781, Jun. 2014.
- [82] I. Walsh, "A strategic path to study IT use through users' IT culture and IT needs: A mixed-method grounded theory," *J. Strategic Inf. Syst.*, vol. 23, no. 2, pp. 146–173, Jun. 2014.
- [83] D. Fletcher and M. Sarkar, "A grounded theory of psychological resilience in olympic champions," *Psychol. Sport Exerc.*, vol. 13, no. 5, pp. 669–678, Sep. 2012.
- [84] B. Wang, Y. Huang, X. Wu, and X. Li, "A fuzzy computing model for identifying polarity of Chinese sentiment words," *Comput. Intell. Neurosci.*, vol. 2015, pp. 1–13, Apr. 2015, doi: 10.1155/2015/525437.
- [85] E. Cambria, S. Poria, D. Hazarika, and K. Kwok, "SenticNet 5: Discovering conceptual primitives for sentiment analysis by means of context embeddings," in *Proc. 32nd AAAI Conf. Artif. Intell.*, 2018, pp. 1795–1820.
- [86] H. Peng and E. Cambria, "CSenticNet: A concept-level resource for sentiment analysis in Chinese language," in *Proc. Process. Comput. Linguistics Intell. Text Process.*, Cham, Switzerland, 2018, pp. 90–104.



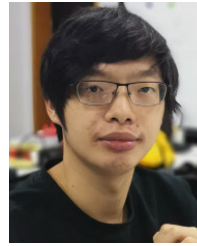
LIANG XU received the B.S. degree in applied psychology and the Ph.D. degree in psychology from Zhejiang University, China, in 2015 and 2020, respectively. He is currently a Postdoctoral Researcher with the Department of Psychology and Behavioral Science, Zhejiang University. His research interests include social network analysis, music emotion recognition, and physiological psychology.



LINJIAN LI received the B.Eng. degree in mechanical engineering from the Zhejiang University of Technology, China, in 2016, and the M.Ed. degree in psychology from Zhejiang University, China, in 2018. His research interests include data science and cognitive psychology.



ZEHUA JIANG received the B.S. degree in psychology from East China Normal University, China, in 2019. She is currently pursuing the Ph.D. degree with the Department of Psychology and Behavioral Science, Zhejiang University.



JIAMING SHI received the B.S. degree in applied psychology from Zhejiang University, where he is currently pursuing the Ph.D. degree with the Department of Psychology and Behavioral Science. His current research interests include the mathematical modeling of human cognition and performance, and transportation safety and human factors in mobile phone usage.



ZAOYI SUN received the B.S. degree in psychology from the Zhejiang University of Technology, China, in 2014, and the Ph.D. degree in psychology from Zhejiang University, China, in 2019. She is currently a Lecturer with the College of Education, Zhejiang University of Technology. Her current research interests include psychological needs, big-data-based behavioral analysis, and user experience.



RUI SUN received the B.S. and master's degrees from Shandong Normal University, China, in 2012 and 2016, respectively. She is currently pursuing the Ph.D. degree with the Department of Psychology and Behavioral Science, Zhejiang University, China. Her current research interest includes user experience.



XIN WEN received the B.S. degree in psychology from Zhejiang University, where she is currently pursuing the Ph.D. degree with the Department of Psychology and Behavioral Science. Her current research interests include cognitive psychology and big-data-based behavioral analysis.



XIUYING QIAN received the B.S. degree from Hangzhou University, China, in 1982, and the Ph.D. degree in experimental engineering psychology from East China Normal University, China, in 1990. She is currently a Professor with the Department of Psychology and Behavioral Science, Zhejiang University, China. Her current research interests include cognitive psychology, unconscious perception, big data analysis, and user experience.

...