



The positive energy of netizens: development and application of fine-grained sentiment lexicon and emotional intensity model

Wenhao Pan¹ · Yingying Han¹ · Jinjin Li² · Emily Zhang³ · Bikai He⁴

Accepted: 10 October 2022 / Published online: 3 November 2022

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2022

Abstract

The outbreak of COVID-19 has led to a global health crisis and caused huge emotional swings. However, the positive emotional expressions, like self-confidence, optimism, and praise, that appear in Chinese social networks are rarely explored by researchers. This study aims to analyze the characteristics of netizens' positive energy expressions and the impact of node events on public emotional expression during the COVID-19 pandemic. First, a total of 6,525,249 Chinese texts posted by Sina Weibo users were randomly selected through textual data cleaning and word segmentation for corpus construction. A fine-grained sentiment lexicon that contained *POSITIVE ENERGY* was built using Word2Vec technology; this lexicon was later used to conduct sentiment category analysis on original posts. Next, through manual labeling and multi-classification machine learning model construction, four mainstream machine learning algorithms were selected to train the emotional intensity model. Finally, the lexicon and optimized emotional intensity model were used to analyze the emotional expressions of Chinese netizens. The results show that *POSITIVE ENERGY* expression accounted for 40.97% during the COVID-19 pandemic. Over the course of time, *POSITIVE ENERGY* emotions were displayed at the highest levels and *SURPRISES* the lowest. The analysis results of the node events showed after the outbreak was confirmed officially, the expressions of *POSITIVE ENERGY* and *FEAR* increased simultaneously. After the initial victory in pandemic prevention and control, the expression of *POSITIVE ENERGY* and *SAD* reached a peak, while the increase of *SAD* was the most prominent. The fine-grained sentiment lexicon, which includes a *POSITIVE ENERGY* category, demonstrated reliable algorithm performance and can be used for sentiment classification of Chinese Internet context. We also found many *POSITIVE ENERGY* expressions in Chinese online social platforms which are proven to be significantly affected by node events of different nature.

Keywords Positive energy · Fine-grained sentiment lexicon · COVID-19 pandemic · Social media analysis

Introduction

From communication to interconnection, from 2 to 5G, the Chinese society has undergone both opportunities and challenges brought by the continuous development of Internet technology. The 49th Statistical Report on China's Internet Development (“Report”) released by the China Internet Network Information Center (CNNIC) in February 2022 shows that up until December 2021, the number of netizens in China has reached 1.03 billion, with an increase of 42.96 million since December 2020 (China Internet Network Information Center, 2022). In this modern era, social media is an important platform for communication and interaction, attracting netizens to participate in the discussion and dissemination of various heated topics during and post-events. As one of the most popular social media platforms in China, Sina Weibo (Weibo) has aggregated huge numbers

Wenhao Pan and Yingying Han contributed equally and should be recognized as co-first authors.

✉ Bikai He
hbikai@126.com

¹ School of Public Administration, South China University of Technology, Guangzhou, China

² School of Psychology, Guizhou Normal University, Guiyang, China

³ Troy High School, Fullerton, CA, USA

⁴ Department of Intelligent Engineering, Guiyang Institute of Information Science and Technology, Guiyang, China

of sadness, fear, anger or happy texts about some hot issues (Sina Weibo Data Center, 2021; Zhang & Yu, 2018). Mining and exploring this kind of massive data from social media is of great significance to the correct and objective understanding of the emotions and attitudes of netizens, and to studying the emotional expression of individuals in the network environment (Han et al., 2022). Meanwhile, it can also provide theoretical reference for government departments to recognize and guide trends related to public opinion evolution. Therefore, the field of mining and utilizing massive text data has been a popular research topic in sociology and psychology.

Emotions are defined as internal mental states that result from an evaluation of events or stimuli (Ekman & Friesen, 1971; Novembre et al., 2019). Therefore, emotions are unstable in a period of time, prone to change with specific events. When the subjective willingness and demands of individuals are consistent with that of the external environment, they will have an optimistic state of mind and display positive emotions (Barbalet, 1998; Novembre et al., 2019). But when the subjective intentions and demands of individuals are inconsistent with that of the external environment, they will have a passive state of mind and negative emotional experience (Novembre et al., 2019). The variability of emotions determines the complexity of its structure and mechanism (Ready et al., 2012; Tavares et al., 2011). In this study, emotional structure is determined by the classification view and dimensional view. Researchers who support the classification view propose that emotions can be divided into different categories, such as negative or positive (Buratto et al., 2014; Kensinger & Corkin, 2004), but this classification fails to consider the granularity of emotions. On this basis, a new emotional classification is proposed by Ekman, in which he defined six standard emotional categories (Ekman & Friesen, 1971; Gao et al., 2020). A large number of later studies have adopted this theory of the six basic emotional expressions (Hutto et al., 2018; Sailunaz & Alhadjj, 2019; Zeng & Zhu, 2019). However, except for the *HAPPY* category that expresses positive emotions, this theory lacks sufficient positive emotional types. In other words, the classification of positive emotions in this theory is not detailed enough. Therefore, the theory of the six basic emotions may not be applicable to the Chinese online community context, which is filled with a large variety of positive emotional texts (Liang et al., 2015; Yang et al., 2019; Zhu & Xu, 2021). In order to more effectively portray positive emotional texts, we categorized texts on the Internet that convey confidence, hope, and encouragement etc. into a new category called *POSITIVE ENERGY*. Another Chinese study adopted a similar approach. They combined Ekman's emotion classification principles and the traditional Chinese emotion classification system to develop an Affective Lexicon Ontology (ALO), which contains a new emotional category *LIKE* (Xu et al.,

2008, 2021). The study suggests that the category *LIKE* is suitable for standard positive emotional texts analysis and can examine emotions more comprehensively.

In addition to emotional classification, emotional intensity should also be regarded as a quantitative dimension that is relatively independent of qualitative differences between emotional categories. For example, emotional expression is not just characterized by classification such as "happiness", "fear" etc., but also by intensity such as "a little happy", "very sad" etc. Some researchers have studied the theoretical support and measurement methods of emotional intensity. Russell (1980) divided human emotions into two dimensions, namely "arousal" (High-low) and "valence" (positive and negative emotions), and many researchers believe that arousal actually reflects emotional intensity (Gong et al., 2018; Kaufman, 1999; Salgado & Kingo, 2019; Sonnemans & Frijda, 1995). Reisenzein (1994) proposed two dimensions, "pleasure" and "arousal" to measure emotional intensity. In the same year, Sonnemans and Frijda (1994) developed a five-dimensional model of emotional intensity. According to these studies, emotional intensity is one of the inherent characteristics of emotional expression, and different types of emotions have different intensity levels.

In the field of emotional psychology, the measurement of emotions mostly adopts two methods: subjective measurement and objective measurement. Among these, the subjective measurement method is generally used to measure emotional expression, using relatively mature and recognized questionnaires with high reliability and validity, such as the Basic Emotion Scale (BES; Power, 2006) and the Pleasure Arousal Dominance Scale (PDAS; Kulviwat et al., 2007). On the other hand, the objective measurement method measures the physiological indicators and behaviors of its subjects, mainly including the measurement of the autonomic nervous system (Moon et al., 2021), the measurement of the activity in the brain (Zhao et al., 2021), and the measurement of behavior (Kümmel & Kimmerle, 2020). All the above-mentioned methods have dominated the field of emotional psychology until simpler and more efficient machine learning (ML) methods appeared and were applied to psychology research. Sentiment analysis is one of the most advantageous and widely used methods of ML, which involves identifying subjective attitudes, emotions, or evaluations of processed texts by using Natural Language Processing (NLP; Medhat et al., 2014). Scholars construct emotional dictionaries and design emotional analysis models according to the emotional characteristics of words to analyze texts representing people's emotional states. It is fair to say that sentiment analysis has been considered an effective supplementary method for emotional psychology research (Crossley et al., 2017).

A suitable sentiment lexicon is the basis for a reliable sentiment analysis. At present, some English sentiment lexicons were widely used in psychology research, such as Linguistic Inquiry Word Count (LIWC; Pennebaker et al.,

2015), General Inquirer (GI; Stone et al., 1966), Affective Norms for English Words (ANEW; Bradley & Lang, 1999) and extended ANEW datasets (Shaikh et al., 2016; War-riner et al., 2013) etc. Among them, ANEW can be the main representative of coarse-grained sentiment lexicon, which contains 1,034 English words with valence, arousal and dominance values (Bradley & Lang, 1999). The LIWC lexicon can be considered as the representative of fine-grained sentiment lexicon (Pennebaker et al., 2015), which contains more than 4,500 English words in more than 88 psychological categories (Lee et al., 2019). Although the research on sentiment analysis of Chinese text started later than in foreign research, the gap between the levels of expertise in this field is small. Some Chinese sentiment lexicons have accumulated an impressive amount of words in recent years, such as HowNet (Dong & Dong, 2003), Simplified Chinese-Linguistic Inquiry and Word Count (SC-LIWC; Cheng et al., 2017), Encoding Emotion in Chinese (Lin & Yao, 2016) and ALO. Among the rest of these lexicons, the ALO lexicon had high coverage and taxonomic classification that were widely used in Chinese sentiment analysis (An et al., 2021; Xu et al., 2008, 2021). However, these sentiment analysis studies have operationalized emotions as the simple factor contrasting emotional texts vs. neutral texts, failing to consider quantifying emotions by combining emotional categories and emotional intensity (Buratto et al., 2014).

At present, research on emotional intensity generally adopts the combination of manual annotation and ML algorithm to construct the intensity model and to analyze the intensity of individual emotional expression by the results of the ML model (Mozetič et al., 2016; Rodríguez & Garza, 2019; Xiong et al., 2019). As early as 1999, Bradley and Lang (1999) had paid attention to the importance of emotional intensity. Combined with the dimensional view of emotion, they used nine-point scales to collect fine-grained ratings for valence, arousal and dominance.¹ Among them, arousal was the dimension of emotional intensity. Meanwhile, based on the manual evaluation method, they also compiled the ANEW lexicon (Bradley & Lang, 1999). Rodríguez and Garza (2019) also used manual annotations to analyze the emotional intensity of a series of texts containing emotions. As a result, they identified three intensities in different emotional categories. Buratto et al. (2014) suggested that researchers should try to classify emotional stimuli as "high", "medium" or "low" intensity to improve the validity of sentiment analysis. Pröllochs et al. (2018) also realized that the polarized emotional lexicon was not sufficient to characterize the emotional expression of users on social media and needed to further distinguish subtle differences

in the intensity of sentiment. They proceeded to produce an automated toolset that combines emotional intensity and categories lexicon (Pröllochs et al., 2018).

Some interesting results have also been obtained by using NLP techniques to study the emotional expression of social events. Bollen et al. (2011) analyzed the relationship between major life events and public sentiment by Twitter users' posts, and it turns out that social events have a significant impact on the intensity of dominant public emotions. Salathé and Khandelwal (2011) investigated the emotional expression of Twitter users during the H1N1 pandemic and found that the emotional intensity of Twitter users showed the characteristics of temporal and spatial changes with the development of the H1N1 pandemic (Salathé & Khandelwal, 2011). In general, while sentiment analysis has become a mainstream research method for studying individual emotional expression (Feng et al., 2020; Kirelli & Arslankaya, 2020), the use of NLP technology to analyze the emotional expression of users in public health emergencies has also been proven feasible (Han et al., 2022; You et al., 2021).

The outbreak of the Coronavirus Disease 2019 (COVID-19) was first reported in Wuhan, China, subsequently, cases of 2019 novel coronavirus pneumonia have been reported worldwide (World Health Organization, 2020). The COVID-19 outbreak has generated significant panic and worry (Han et al., 2022; Nicomedes & Avila, 2020). In some ways, the COVID-19 pandemic broadly holds the world in its grip (Blanken et al., 2021). As this pandemic is characterized by suddenness, unpredictability and severe individual and public damage, it has an immeasurable negative impact on the emotional expression and behavioral performance of individuals (Havnen et al., 2020). Public emotions may even change dramatically because of the sudden changes in their surroundings (Emodi-Perlman et al., 2021). In addition, due to the strict lockdown policies implemented by various countries, people who were locked down at home will lack sufficient information sources and real-world social support (Rubin & Wessely, 2020; Wang et al., 2020). These strict policies might lead to repeated perceptions of various emotional experiences brought about by the pandemic information. Such a single emotional experience will not only overconsume the cognitive resources used for everyday life but also harm the physical and mental health of individuals (Emodi-Perlman et al., 2021; Sønderskov et al., 2020). Therefore, it is necessary to explore the emotional expression characteristics of netizens during the COVID-19 pandemic.

In summary, previous studies have provided a practical way to research the emotional expression of netizens during the COVID-19 outbreak. Based on the existing research, this study intends to construct the objective and general fine-grained sentiment lexicon and emotional intensity model of Weibo netizens, which would fill the gaps in previous research of the Chinese online context. In addition, the above

¹ <https://airtable.com/shrnVoUZrWu6riP9b/tblf3VIxnvNiApt7M/viwOVb3xoVejnmhvG>

sentiment analysis tools constructed in this study were verified and applied in the emotional expression analysis of Chinese netizens during the COVID-19 outbreak. This research could play a certain role in easing the negative emotions and increasing positive emotions of Chinese netizens.

Measurement instruments

The construction of original Weibo corpus

A general and accurate corpus is the foundation of an authentic fine-grained sentiment lexicon construction (Feng et al., 2020). Because there is no existing public corpus of Chinese social media original texts, we constructed a Chinese corpus for our analyses. In this study, Sina Weibo² was selected as the data source of the research corpus, while Python was used to compile the crawler of Weibo text, and the UIDs of Weibo users were randomly selected as clues to collect. The Weibo corpus we collected contains all original Weibo texts published by the users since registration. The privacy of Weibo users was strictly protected in the whole process in adherence to the ethical principles reference listed by Kosinski et al. (2015). Furthermore, this study is also approved by the Ethics Committee of the School of Psychology, Guizhou Normal University (Ethics Approval No. GZNUPSY53). The details of corpus construction are as follows.

Original Weibo texts collection

During the period from October 2021 to November 2021, we randomly captured the public information of followers of official media accounts in Weibo including "CCTV News" and "Xinhua News Agency". At the same time, we combined UID to collect the original texts posted by these randomly selected Weibo users as the corpus. After excluding users with less than 50 original posts on Weibo (Xu et al., 2021), 7,415,252 original Weibo texts for corpus construction were finally obtained.

Textual data cleaning

The original Weibo texts contain an immense amount of personalized texts, which are potentially inconducive for emotional expression analysis. Thus, in order to improve the accuracy of sentiment analysis, the original texts were cleaned prior to analysis. The specific cleaning process is as follows: 1) Removing duplicate texts based on URL. 2) Excluding the texts that netizens forwarded. 3) Eliminating

invalid contents by regular matching techniques, such as English vocabulary, utf8mb4 emojis, hashtags, extra spaces, and URL links in the texts. After data cleaning, a total of 6,525,249 Chinese text data were retained.

Chinese word segmentation

Generally speaking, sentiment analysis relies upon words and their associated meaning (Havey, 2020). But an obstacle posed by Chinese texts is that Chinese sentences are represented as vocabularies or character strings without natural delimiters. Chinese word segmentation technology is used to identify word sequences in a sentence and mark (usually SPACE) boundaries in appropriate places (Shi et al., 2018). In this study, we availed the accurate mode of Jieba,³ a popular open-sourced Python module for Chinese text segmentation. After textual data cleaning and Chinese word segmentation, the original texts were finally usable as a corpus for the lexicon construction.

The construction of Weibo fine-grained sentiment lexicon

Extraction of seed words

Previous studies mostly adopted coarse-grained sentiment analysis which is based on the emotional positivity or negativity of vocabularies (Leis et al., 2019). However, there are relatively fewer studies on the fine-grained classification of emotions, and the existing researches lack a consistent classification standard of emotion. This study is an extension of the prior research. In the research of Chinese NLP, the ALO lexicon is a widely used fine-grained sentiment lexicon which includes the six categories of emotion words (*HAPPY*, *ANGRY*, *SAD*, *FEAR*, *DISGUST*, and *SURPRISE*) from the Ekman emotion classification principle and a *LIKE* category from the Chinese classical emotion classification system.

However, the source of the ALO lexicon is mainly standardized dictionaries, such as "Modern Chinese Classification Dictionary", "Chinese Adjective Dictionary" etc. (Xu et al., 2008, 2021), so the ALO lexicon may not be directly applicable in the Internet context where language is relatively liberalized and personalized. Therefore, it is necessary to construct the domain lexicon according to the needs of the research and select the seed words from the normalized dictionary, to construct a seed word set. In this study, the typical emotional vocabularies involved from the ALO lexicon can be used as the seed word of the domain lexicon in the network context.

² <https://weibo.com>

³ <https://github.com/fxsjy/jieba>

Table 1 Example of emotion seed words in Weibo

Category	Seed words	Count
HAPPY	开心 ("happy"), 快乐 ("joy"), 微笑 ("smiling"), 捧腹大笑 ("laughing")	89
ANGRY	愤怒 ("anger"), 恼怒 ("irritation"), 狂怒 ("rage"), 愤愤不平 ("indignation")	83
SAD	悲痛 ("grief"), 绝望 ("despair"), 孤独 ("alone"), 哭泣 ("crying")	97
FEAR	惊慌 ("panic"), 畏惧 ("fear"), 胆怯 ("timidity"), 焦虑 ("anxiety")	92
DISGUST	憎恶 ("aversion"), 自负 ("vanity"), 丢脸 ("reproach"), 发牢骚 ("ranting")	94
SURPRISE	惊讶 ("wonder"), 奇迹 ("marvel"), 吃惊 ("flabbergasted"), 惊愕 ("jaw-dropping")	42
POSITIVE ENERGY	祝福 ("blessing"), 信任 ("trust"), 可靠 ("reliability"), 钦佩 ("admiration")	82

Referencing previous studies (Luo et al., 2015; Xu et al., 2021), this study recruited three volunteers (including a PhD student and two master students in psychology) to manually select 100 words that best describe the state of each category of emotional vocabulary in the ALO lexicon. For the unanimously selected words, these words were directly included into the seed word set of this category of emotion. For the inconsistent words, each word was discussed by these three volunteers to reach a consensus. A total of 579 emotion seed words were included in this study, and the examples of seed words are shown in Table 1.

The composition of fine-grained sentiment lexicon

In this study, the vocabulary source of fine-grained sentiment lexicon is the original Weibo corpus mentioned above. Specifically, the Word2Vec model was used to conduct similar word expansion training, and the emotional expression words with the highest similarity scores to seed words were obtained. Word2Vec is a mainstream word embedding approach in the NLP research field and can be used to build a sentiment dictionary (Roh et al., 2019; Bai et al., 2014). Word2Vec technology can be trained on a large amount of data (Zhu et al., 2017) so as to obtain the words that are highly similar and correlate to seed words and use cosine value as similarity index (Duong et al., 2018).

In this study, the original Weibo corpus was adopted in the Word2Vec model training process. The details of this process are as follows: First, the continuous bag of words (CBOW) function of the Word2Vec toolkit was used to train the appropriate model and calculated the word vector results of words in the corpus. Second, the seed words were imported into the trained model, and calculated the top 50 emotional expression words with the highest similarity score ($\cosine \geq 0.7$) to the seed words (Xu et al., 2021). In order to get a higher Word2Vec model effect, we performed Word2Vec model on corpora of different orders of magnitude (including 500,000 words, 1,000,000 words, 2,000,000 words, and 6,000,000 words) respectively. We took the similarity level (\cosine values) of the specified word as the specific evaluation index and selected a corpus with a higher cosine value. It was found that the corpus constructed with 6

million words could obtain a higher Word2Vec model effect. Finally, three volunteers (including a PhD student and two master students in psychology) were recruited to perform manual word filtering in order to eliminate ambiguous words and uncommon words in the Chinese online context. Similarly, each word or phrase was discussed by these volunteers in order to reach a consensus. After these operations, the fine-grained sentiment lexicon was finally built. The vocabulary statistics of the lexicon are shown as follows: 459 *HAPPY* words, 381 *ANGRY* words, 382 *SAD* words, 442 *FEAR* words, 716 *DISGUST* words, 126 *SURPRISE* words, and 512 *POSITIVE ENERGY* words.

The verification of fine-grained sentiment lexicon

In order to gauge the validity of the fine-grained sentiment lexicon, this study employed coverage rate and classification accuracy to verify the validity of the above lexicon and compared it with those of the ALO lexicon. Among them, coverage rate refers to the proportion of the number of Weibo texts containing emotional words in the lexicon to the total number of texts. Classification accuracy is a common indicator of tool performance judgment, including precision rate, recall rate, and F_1 -score.

The specific process of lexicon verification is as follows: First, 1,800 Weibo posts were randomly selected as the test dataset, and 3 volunteers were recruited to label the sentiment categories. In addition, since the same Weibo text may express multiple emotions, we set the principle that a text can be labeled as multiple sentiment categories.

Second, the ALO lexicon and fine-grained sentiment lexicon were loaded separately and performed vocabulary matching on the text of the test set. The results showed that compared with the ALO lexicon (with the coverage rate of 72.6%), our fine-grained sentiment lexicon had a higher coverage rate (with the coverage rate of reaching 84.3%).

Finally, we analyze the classification accuracy of the fine-grained sentiment lexicon. The accuracy indicators mainly include precision rate, recall rate, and F_1 -score, where the F_1 -score is the weighted harmonic mean of the precision rate and recall rate. In addition, the subset

Table 2 Comparison of lexicon accuracy results

Category	Fine-grained sentiment lexicon			ALO lexicon		
	Precision	Recall	F_1	Precision	Recall	F_1
HAPPY	0.68	0.71	0.69	0.54	0.57	0.55
ANGRY	0.73	0.62	0.67	0.42	0.27	0.32
SAD	0.69	0.58	0.63	0.63	0.47	0.53
FEAR	0.58	0.72	0.64	0.49	0.36	0.41
DISGUST	0.64	0.66	0.65	0.36	0.55	0.43
SURPRISE	0.43	0.64	0.51	0.49	0.36	0.41
POSITIVE ENERGY	0.56	0.64	0.59	0.47	0.39	0.42

of negative words was imported into the text sentiment classification, so as to carry out more accurate results. The results showed that the fine-grained sentiment lexicon had a higher classification accuracy. The results are shown in Table 2.

The construction of Weibo emotional intensity model

Emotional expression is not only including the characteristics of emotional categories, but also includes the difference in emotional intensity. By comprehensively analyzing the characteristics of netizens' emotional categories and emotional intensity, we can better investigate individuals' emotional attitudes towards their own life and specific life events. Therefore, this study constructed the supervised ML model suitable for the emotional intensity based on the fine-grained sentiment lexicon.

Manual labeling of the training dataset

The core of supervised ML algorithms for text analysis is to take the manually annotated training data as the criterion of ML model training. In this part, we selected the above-mentioned test textual dataset (in “[The verification of fine-grained sentiment lexicon](#)” section) that had been classified into emotional categories and recruited volunteers for manual labeling of emotional intensity.

The specific process was as follows: First, calculate the consistency of the scores of different raters. Kendall's coefficient of concordance was used to determine the consistency of scoring in this study. By obtaining Kendall's coefficient of concordance, a good evaluation result or a good scorer could be selected objectively (Ferrari et al., 2009; Guo et al., 2020). In order to get more accurate results, 200 texts from different emotional categories of the test dataset were randomly selected, and five raters (including three master students in psychology and two undergraduate students in Chinese language

and literature) independently rated the emotional intensity levels from 1 to 5. Subsequently, Kendall's coefficient of concordance (W) of the emotional intensity scores was calculated. The results showed that there was a high level of consistency among different raters ($W_{HAPPY}=0.79$, $W_{ANGRY}=0.82$, $W_{SAD}=0.82$, $W_{FEAR}=0.79$, $W_{DISGUST}=0.77$, $W_{SURPRISE}=0.74$, $W_{POSITIVE ENERGY}=0.79$, $P_s < 0.01$). Then, these raters were asked to continue evaluating the texts of the seven emotion categories, and 2500 sample sets were finally obtained for subsequent emotional intensity model training.

The training of emotional intensity models

The protocol of this study is structured following the previous ML studies' recommendations. First comes text vectorization and dimension reduction operation. In this study, we used vocabulary as the feature of the emotional intensity model, that was, the Chinese vocabulary contained in each text was used as the classification basis. In order to avoid too complex calculations, dimensionality reduction operations on text features were required (Hoke et al., 2020). In this study, we independently adopted the term frequency-inverse document frequency technology (TF-IDF) and set the high threshold. Then, the dimensionality reduction method was used for these Weibo texts. After the TF-IDF operation, the remaining word vectors were included in the emotional intensity prediction task.

The second step is ML algorithm selection operation. This study used the supervised ML algorithms to analyze the emotional intensity, with the purpose of constructing a five-level classification model of the emotional intensity of these texts (Sasaki et al., 2009). First, according to the recommendations of previous studies (Muraoka et al., 2019), the dataset of 2500 texts was divided into a training set of 2000 texts (400 texts of level 5 emotional intensity) and a test set of 500 texts. Subsequently, the widely used ML algorithms were selected into training a multi-classification model of emotion intensity, mainly

Table 3 Emotional intensity model performance measures

Category	Algorithm	Original performances			After hyperparameter tuning		
		Precision	Recall	F_1	Precision	Recall	F_1
HAPPY	NB	0.67	0.68	0.67	-	-	-
	RF	0.89	0.89	0.89	0.90	0.89	0.89
	KNN	0.64	0.64	0.63	-	-	-
	LR	0.84	0.84	0.84	-	-	-
ANGER	NB	0.64	0.62	0.62	-	-	-
	RF	0.87	0.87	0.87	0.87	0.88	0.87
	KNN	0.56	0.56	0.55	-	-	-
	LR	0.83	0.84	0.84	-	-	-
SAD	NB	0.63	0.61	0.62	-	-	-
	RF	0.84	0.83	0.84	0.85	0.84	0.84
	KNN	0.59	0.59	0.59	-	-	-
	LR	0.79	0.79	0.79	-	-	-
FEAR	NB	0.68	0.62	0.62	-	-	-
	RF	0.89	0.90	0.89	0.90	0.91	0.91
	KNN	0.59	0.58	0.58	-	-	-
	LR	0.82	0.81	0.82	-	-	-
DISGUST	NB	0.68	0.67	0.67	-	-	-
	RF	0.88	0.88	0.88	0.89	0.88	0.89
	KNN	0.56	0.55	0.55	-	-	-
	LR	0.80	0.79	0.79	-	-	-
SURPRISE	NB	0.68	0.66	0.66	-	-	-
	RF	0.86	0.87	0.86	0.87	0.87	0.87
	KNN	0.56	0.55	0.55	-	-	-
	LR	0.81	0.81	0.81	-	-	-
POSITIVE ENERGY	NB	0.74	0.73	0.73	-	-	-
	RF	0.88	0.87	0.88	0.89	0.87	0.88
	KNN	0.63	0.62	0.62	-	-	-
	LR	0.81	0.82	0.81	-	-	-

NB refers to Naive Bayesian algorithm, RF refers to Random Forest algorithm, KNN refers to K-Nearest Neighbor algorithm, and LR refers to Logistic Regression algorithm

including Naive Bayesian (NB), Random Forest (RF), K-Nearest Neighbor (KNN) and Logistic Regression (LR; Keshmiri et al., 2020).

The third step is emotional intensity model training. The fine-grained sentiment lexicon and the four ML algorithms were imported to conduct model training of different sentiment categories in the training set. At the same time, the manual labeling text data of emotional intensity were used as the criterion to analyze the accuracy of the ML model. It turned out that, on the whole, the precision rate, recall rate and F_1 -score of the RF algorithm were the best with the default parameter value among these algorithms. The original performances are summarized in Table 3.

The fourth step is algorithm hyperparameter tuning and prediction operation. After comparing different ML

algorithms, the RF showed the highest algorithm performance in the task of emotional intensity classification. Therefore, the RF algorithm was selected as the prediction algorithm. Meanwhile, to provide optimal prediction performances, hyperparameter tuning technology was performed with the GridSearchCV estimator. The range of GridSearchCV estimator were set as follows: " $N_estimators$ ": range (50, 200, 10), " max_depth ": range (2, 10, 2), " $max_features$ ": range (3, 11, 2), " $min_samples_leaf$ ": range (10, 100, 10). The results showed that the optimized RF had better algorithm performance after hyperparameter tuning. And the optimal parameters were $n_estimators = 80$, $max_depth = 6$, $max_features = 7$, $min_samples_leaf = 30$, $random_state = 10$. The performances after hyperparameter tuning are also shown in Table 3.

Table 4 Descriptive statistics of sentiment category and intensity

Category	Sentiment category statistics		Emotional intensity level statistics (<i>n</i>)					
	Counts	Proportion (%)	Level 1	Level 2	Level 3	Level 4	Level 5	<i>M</i> ± <i>SD</i>
HAPPY	182,815	20.41	43,024	29,134	18,040	54,779	37,838	3.08 ± 1.49
ANGRY	94,787	10.59	11,805	12,430	17,777	21,739	31,035	3.50 ± 1.38
SAD	122,144	13.63	22,077	13,794	28,991	9,136	48,147	3.39 ± 1.53
FEAR	210,102	23.46	11,635	40,477	79,053	42,892	36,045	3.24 ± 1.12
DISGUST	185,406	20.7	30,672	14,254	79,982	36,910	23,588	3.05 ± 1.20
SURPRISE	38,936	4.34	4,573	12,230	8,790	11,011	2,332	2.85 ± 1.13
POSITIVE ENERGY	366,872	40.97	10,196	30,245	184,361	91,803	50,267	3.39 ± 0.92
With Sentiment	895,505	52.46	-	-	-	-	-	-
Without Sentiment	811,436	47.54	-	-	-	-	-	-

The application of Weibo fine-grained sentiment lexicon and emotional intensity models

In this research, the Weibo fine-grained sentiment lexicon and the emotional intensity model were constructed. To demonstrate possible application scenarios of these tools, taking COVID-19 as a case, the lexicon and intensity models were imported to text analysis on the emotional expression of netizens.

Research method

This study explored the emotional expression characteristics of Weibo netizens during the COVID-19 pandemic through the following three steps. First, acquire the original Weibo texts during the COVID-19 pandemic. Referring to the description of the COVID-19 pandemic in Mainland China in the white paper "Fighting COVID-19: China in Action" (The State Council Information Office of China, 2020), we set the time period in our study as from December 27, 2019, to April 13, 2020. Subsequently, we obtained a total of 1,808,671 original Weibo texts with keywords, such as "COVID-19 pandemic" and "new crown pneumonia". The data search strategy terms were presented in Appendix Table 7. And we used the text preprocessing method for data cleaning and word segmenting (in "The construction of original Weibo corpus" section). Finally, a total of 1,706,941 Weibo texts were used.

Secondly, the Weibo fine-grained sentiment lexicon and emotional intensity models were used to analyze the characteristics of netizens' emotional expressions. The Weibo fine-grained sentiment lexicon was imported to distinguish the sentiment categories of all Weibo texts. At the same time, the emotion intensity model was called to predict the intensity in different emotion categories. Subsequently, we contrasted the emotional expression

of netizens on the same emotional intensity during the COVID-19 pandemic. To achieve this, this study used the emotional intensity of each emotion category on the day as "The total scores of Weibo texts on a certain emotion of this day divided by the total number of Weibo texts on this day" (Ye et al., 2016). Furthermore, we took 24 h as the time unit and made the trend graph of emotional intensity over time.

Finally, the influence of social events on emotional expression during the COVID-19 pandemic was analyzed. Based on the inflection point of the emotional expression in the trend graph, we explored the impact of social events that occurred during the COVID-19 pandemic on the emotional expression of netizens. In this study, an independent sample T-test was performed on the emotional intensity scores of the day before and after each inflection point corresponding to the social event. Meanwhile, to determine which emotion is dominant after the social event occurred, a 2 × 3 (time * emotions) multivariate analysis of variance (MANOVA) was conducted.

Research results

The overall emotional expression of netizens during the COVID-19 pandemic

This study loaded the Weibo fine-grained sentiment lexicon for sentiment analysis. The results showed that Weibo texts with sentiment accounted for 52.46%, and among them, *HAPPY* emotion accounted for 20.41%, *ANGRY* emotion accounted for 10.59%, *SAD* emotion accounted for 13.63%, *FEAR* emotion accounted for 23.46%, *DISGUST* emotion accounted for 20.70%, *SURPRISE* emotion accounted for 4.34%, and *POSITIVE ENERGY* emotion accounted for 40.97%. In addition, Weibo texts without sentiment accounted for 47.54%.

The emotional intensity of *ANGRY* was the highest (3.50 ± 1.38), followed by *SAD* (3.39 ± 1.53) and

POSITIVE ENERGY (3.39 ± 0.92), and the intensity of *SURPRISE* was the lowest (2.85 ± 1.13). In addition, the intensity of *HAPPY* (3.08 ± 1.49), *FEAR* (3.24 ± 1.12), and *DISGUST* (3.05 ± 1.20) was at the medium level. The specific analysis results are shown in Table 4.

The overall trends of netizens' emotional expression during the COVID-19 pandemic

Taking 24 h as the time unit, the overall trends of the emotional expression were analyzed. It turned out that the intensity of *HAPPY* in the early stage of the COVID-19 pandemic was the highest among various emotions. However, with the spread of the COVID-19 pandemic, the intensity of *HAPPY* had dropped sharply since January 19, 2020, and it had maintained a relatively stable low level. In addition, the intensity of *POSITIVE ENERGY* which represents hope, belief, and admiration had shown a clear growth trend since January 19, 2020, and had maintained a relatively high level ever since. The intensity of *SAD* remained relatively stable throughout the COVID-19, but there was a clear peak of *SAD* around April 4, 2020. The intensity of *SURPRISE* and *ANGRY* was at a relatively low level from December 27, 2019, to April 13, 2020. Furthermore, the intensity of *DISGUST* was at a moderate level throughout the pandemic. The specific trends of emotional intensity are shown in Fig. 1.

The impact of key social events on emotional expression

With reference to the white paper "*Fighting COVID-19: China in Action*", we selected two Turning Points events during the COVID-19 pandemic as the important nodes, which including critical events in the stage of "*Swift Response to the Public Health Emergency*" and "*Initial Victory in a Critical Battle*" (The State Council Information Office of China, 2020). The first node event was that "*The NHC assembled a high-level national team of senior medical and disease control experts, and the team determined that the new coronavirus was spreading between humans*", which took place on January 19, 2020 (Fig. 1). The second event was that "*A nationwide ceremony was held on the traditional Tomb-sweeping Day to pay tribute to all those who had given their lives in fighting Covid-19, and others who had died of the novel coronavirus*", which took place on April 4, 2020 (Fig. 1). By analyzing the changes in the emotional expression of Weibo users before and after these events, we explored the impact of the key "Turning Points" events during the COVID-19 pandemic on Weibo users' emotional expression.

First, we analyzed Event 1 (January 19, 2020). The independent-sample t-test results showed that the emotional intensity of *FEAR*, *ANGRY* and *POSITIVE ENERGY* were significantly different before and after Event 1

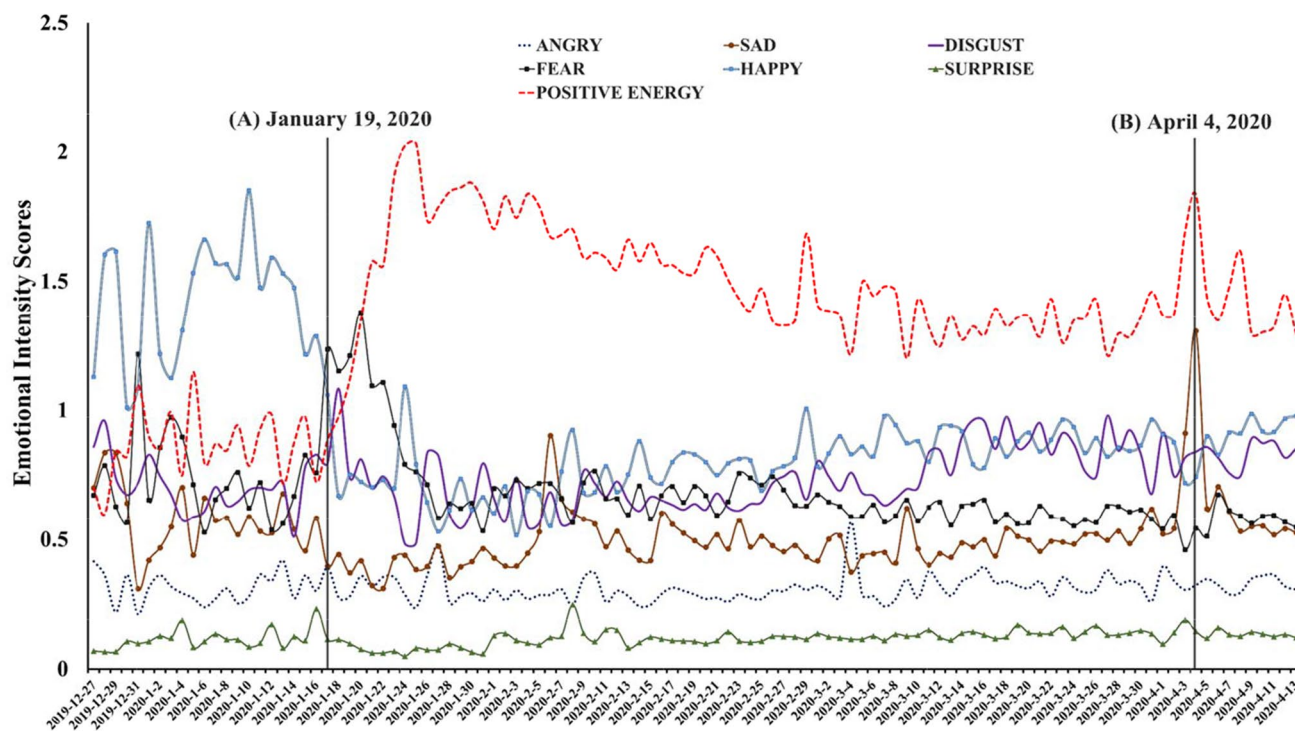


Fig. 1 The trend graph of emotional intensity regarding the COVID-19 pandemic from December 27, 2019, to April 13, 2020. The values on the vertical axis reflect the level of emotional intensity. (A) Node

event in the stage of "Swift Response to the Public Health Emergency" on January 19, 2020; (B) Node event in the stage of "Initial Victory in a Critical Battle" on April 4, 2020

Table 5 Results of difference in emotional intensity of node events

Category	Time point	Event 1				Event 2			
		Count	Mean score	<i>SD</i>	<i>t</i>	Count	Mean score	<i>SD</i>	<i>t</i>
ANGRY	Before	1419	0.28	0.90	-2.52*	3785	0.31	0.91	-1.02
	After	2608	0.36	1.03		8520	0.32	0.92	
DISGUST	Before	1419	0.74	1.48	-0.15	3785	0.82	1.65	-0.67
	After	2608	0.81	1.53		8520	0.84	1.66	
FEAR	Before	1419	1.21	1.69	-2.98**	3785	0.46	1.08	-3.87***
	After	2608	1.38	1.74		8520	0.54	1.14	
HAPPY	Before	1419	0.75	1.64	0.51	3785	0.72	1.58	-0.79
	After	2608	0.73	1.50		8520	0.74	1.56	
SAD	Before	1419	0.37	1.22	-1.14	3785	0.91	1.69	-11.57***
	After	2608	0.42	1.29		8520	1.31	1.88	
SURPRISE	Before	1419	0.10	0.54	1.30	3785	0.19	0.78	2.93**
	After	2608	0.08	0.49		8520	0.15	0.69	
POSITIVE ENERGY	Before	1419	1.12	1.58	-4.33***	3785	1.70	1.89	-3.77***
	After	2608	1.35	1.69		8520	1.83	1.88	

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$ **Table 6** Pairwise interaction between emotions and time point of node events

Node events	Interaction term	Comparison of emotions	<i>F</i>	<i>p</i>
Event 1	Emotions × Time Point (With <i>FEAR</i> as a contrast)	<i>FEAR</i> vs. <i>ANGRY</i>	85.638	<0.001
		<i>FEAR</i> vs. <i>DISGUST</i>	31.369	<0.001
		<i>FEAR</i> vs. <i>HAPPY</i>	52.982	<0.001
		<i>FEAR</i> vs. <i>SAD</i>	82.472	<0.001
		<i>FEAR</i> vs. <i>SURPRISE</i>	157.979	<0.001
		<i>FEAR</i> vs. <i>POSITIVE ENERGY</i>	0.150	0.699
	Emotions × Time Point (With <i>POSITIVE ENERGY</i> as a contrast)	<i>POSITIVE ENERGY</i> vs. <i>ANGRY</i>	459.223	<0.001
		<i>POSITIVE ENERGY</i> vs. <i>DISGUST</i>	38.793	<0.001
		<i>POSITIVE ENERGY</i> vs. <i>FEAR</i>	0.162	0.687
		<i>POSITIVE ENERGY</i> vs. <i>HAPPY</i>	63.586	<0.001
		<i>POSITIVE ENERGY</i> vs. <i>SAD</i>	97.005	<0.001
		<i>POSITIVE ENERGY</i> vs. <i>SURPRISE</i>	181.628	<0.001
Event 2	Emotions × Time Point (With <i>SAD</i> as a contrast)	<i>SAD</i> vs. <i>ANGRY</i>	68.415	0.000
		<i>SAD</i> vs. <i>DISGUST</i>	67.232	<0.001
		<i>SAD</i> vs. <i>FEAR</i>	46.935	<0.001
		<i>SAD</i> vs. <i>HAPPY</i>	66.289	<0.001
		<i>SAD</i> vs. <i>SURPRISE</i>	92.578	<0.001
		<i>SAD</i> vs. <i>POSITIVE ENERGY</i>	31.537	<0.001

($t_{FEAR} = -2.982$, $P_{FEAR} < 0.01$; $t_{ANGER} = -2.516$, $P_{ANGER} < 0.05$; $t_{POSITIVE ENERGY} = -4.330$, $P_{POSITIVE ENERGY} < 0.001$). However, other emotion categories did not reach the significance level ($P_s > 0.05$), and specific results of Even 1 are presented in Table 5.

The results of MANOVA showed that the comparison of the emotional intensity of *FEAR* and *POSITIVE ENERGY* had no significant interaction with time point [$F(1, 31,284) = 0.150$, $P = 0.699$], which proves that there

is no significant difference in the amount of change over time. However, the pairwise comparison of the emotional intensity increments between these two emotions (*FEAR* and *POSITIVE ENERGY*) and the remaining five emotions has significant interaction with the time point ($P_s < 0.001$). The detailed results are presented in Table 6. The results of Event 1 indicate that emotional intensity levels of the *FEAR* and *POSITIVE ENERGY* are significantly different from the other five emotions over time. At the same time,

after the occurrence of "*the high-level national team determined that the new coronavirus was spreading between humans*", the emotional increments of *FEAR* and *POSITIVE ENERGY* of Weibo users are higher than the emotional increments of other emotions.

Second, the analysis of the Event 2 (April 4, 2020). The independent-sample t-test results of the Weibo text data before and after Event 2 show that the emotional intensity of *FEAR*, *SAD* and *POSITIVE ENERGY* increased significantly ($t_{\text{FEAR}} = -3.871$, $t_{\text{SAD}} = -11.570$, $t_{\text{POSITIVE ENERGY}} = -3.767$, $P_s < 0.001$), and the emotional intensity of *SURPRISE* reduced significantly ($t = 1.962$, $P < 0.01$). Furthermore, the independent sample t-tests of other emotional intensity did not reach the significance level ($P_s > 0.05$), and specific results of Event 2 are presented in Table 5.

The results of MANOVA showed that the comparison of the emotional intensity increments of *SAD* and other emotional intensity had a significant interaction with time points, and specific results of Event 2 are shown in Table 6. The results indicate that the emotional intensity of the *SAD* was significantly different from the other six emotions over time. At the same time, after the occurrence of "*the nationwide ceremony was held on the traditional Tomb-sweeping Day*", the emotional increments of Weibo users' *SAD* were higher than the emotional increments of other emotions.

General discussion

Discussion of research tools

This study aims to investigate the emotional expression characteristics of netizens during the COVID-19 pandemic, and analyze the positive energy expression in Chinese social media platforms. To achieve these objectives, we built a fine-grained sentiment lexicon and Weibo text emotional intensity models that are suitable for the Chinese online network language environment.

After the test of performance indicators, we found that the self-built research tools could comprehensively investigate the emotional expression characteristics of netizens. Our study also provided a new perspective on the individual psychological dynamic analysis by applying the standardizing work processes in the ML field. We acknowledged the fact that most of the lexicons constructed by previous studies have shortcomings, such as the polarization of emotional classification and excessive negative emotional categories, which is why their results may have difficulty covering the positive emotional expression in the Chinese online language

environment. In order to investigate the emotional expression characteristics of netizens, especially in the context of the pandemic, this research innovatively incorporated the *POSITIVE ENERGY* emotional category. And after examination, it was found that Chinese netizens do express more positive energy emotion on social media platforms.

Additionally, this study further constructed an ML multi-classification model that is suitable for analyzing emotional intensity. This emotional intensity model (Based on the RF algorithm) had better precision scores, recall scores and F_1 scores of the Weibo netizens' emotional intensity prediction model under different emotion categories. After applying the model, we found that the intensity of positive energy expression among most netizens in China continued to maintain a high level during the COVID-19 outbreak. Consistent with previous studies (Naslund et al., 2020; Ostic et al., 2021; Ranieri et al., 2021), our study also evidenced salient findings that increased use of online social media can effectively promote positive psychological state and positive emotional expression.

Furthermore, it is difficult to get a complete result by using emotional dictionaries or intensity models alone if we need to analyze the characteristics of Weibo users' emotional expressions. Bandhakavi et al. (2021) pointed out that when studying emotion expression, researchers should consider not only the categories of emotion but also the intensity of emotion. Given that previous studies have rarely considered the categories and intensities of emotions in a comprehensive manner, this study addresses this shortcoming by constructing two research tools for measuring netizens' emotional expression with high reliability and validity for the domestic study of Chinese text analysis.

Discussion on netizens' emotional expression during the COVID-19 pandemic

According to the emotional expression results, we graphed the trend of emotional expression as shown in Fig. 1. Specifically, we classified the sentiment of 1,706,941 original Weibo texts during the COVID-19 pandemic. The results show that posts with emotion accounted for 52.46% and without emotion accounted for 47.54%, which is consistent with the previous studies (Raghupathi et al., 2020; Zeng & Zhu, 2019). In the subsequent analysis process, only the emotional texts were analyzed. Below, we illustrated the impact of the pandemic on the netizens' emotional expression from the overall trend and the trend before and after key node events.

The overall trend of netizens' emotional expression during the COVID-19 pandemic

From the time course of emotional expression, we found that before the COVID-19 outbreak, a large variety of texts in which netizens shared their lives and emotional states existed on social media platforms, with the content of these emotional texts varies from person to person. Therefore, there might be irregular fluctuations in each category of emotions. However, around January 19, 2020, there was a key emotional expression turning point. The emotional category that had changed most obviously was *HAPPY*. Since the time node of January 19, *HAPPY* emotion had shown a clear downward trend and maintained at a relatively low level. For the Chinese people, January 19 should be an extremely festive day, as it was the last week before the Chinese Lunar New Year (Chinese Spring Festival), when all Chinese people lived in a festive atmosphere. However, after the news of the "*high-level national team of senior medical and disease control experts determined that the new coronavirus was spreading between humans*" was broadcast, the joy of all Chinese citizens ended abruptly.

It is noteworthy that after January 19, *POSITIVE ENERGY* emotion increased rapidly, reached a peak on January 24, and maintained a high level throughout the COVID-19 pandemic. This phenomenon might be correlated to China's rapid and effective pandemic prevention and control measures. Starting from January 24, 346 national medical teams and a large number of medical staff from various provinces had been mobilized to urgently assist Wuhan, Hubei province (The State Council Information Office of China, 2020). When the public believed that the pandemic was preventable and controllable, they had more confidence in overcoming the pandemic, which might lead to the highest *POSITIVE ENERGY* emotional expressed by netizens (Twenge & Joiner, 2020). Soon, the negative emotions that were common at the beginning of the COVID-19 pandemic gradually decreased, and positive emotions (i.e., hope, confidence, encouragement) of netizens gradually increased (Twenge & Joiner, 2020).

Judging from the overall characteristics of different emotional expressions, the intensity of expression of *SURPRISE* was relatively low and stable among various emotions during the COVID-19 outbreak stage and the pandemic prevention and control stage, which is consistent with previous studies (Medford et al., 2020). Ye et al. (2020) used text analysis technology to investigate the emotional level of netizens towards COVID-19, and found that when faced with negative life events, netizens' expression of surprise emotions will be at a stable low level. Surprise is an emotion that individuals

might feel only something unexpected happens right next to themselves (Aslam et al., 2020). Since all Chinese nationals were asked or required to accept the home quarantine order, which curbed the continuous spread of the pandemic, the coronavirus did not directly appear in the immediate surrounding of most Chinese individuals, explaining the common low level of surprise among domestic individuals. Min et al. (2021) used ML to investigate the emotional responses of Twitter users during the COVID-19 pandemic, reaching the same conclusion that the trajectory of surprise did not show significant relative changes and maintained at a low level. The results of this study also further illustrated that the individuals' surprise emotion might be difficult to be affected by persistent negative life events (King et al., 2008).

Based on pandemic critical nodes to analyze netizens' various emotional expressions

Except for the *SURPRISE* emotion, we found that the overall trend of other emotions was not obvious. Therefore, we analyzed other emotion categories and found that after the specific key node event occurred, the emotional expression of netizens had shown an obvious turning point. Based on the results of the multivariate analysis of variance, we further analyzed the node events that have sentiment inflection points.

Specifically, after the occurrence of Event 1, the expression intensity of *FEAR*, *POSITIVE ENERGY*, and *ANGRY* of Weibo users increased significantly. In terms of *FEAR* emotion, before the news of "*human-to-human transmission was confirmed*", the information related to COVID-19 that the public knew was "*pneumonia patients of unknown cause in Wuhan*" (December 27, 2019) and "*no clear evidence of human-to-human transmission*" (January 14, 2020). Therefore, these news messages might not have impact on people's living conditions and increased their *FEAR*. However, for the public, after the official announcement of "*human-to-human transmission of new crown pneumonia*" from NHC of China, the novel coronavirus pneumonia turned uncontrollable. Due to the sudden and unpredictable nature of public health emergencies, the public started to feel powerless to do anything about the development of the pandemic. They might also have been unable to adapt to the sudden changes in their living environment, which could cause tension, fear and other negative emotions, with the *FEAR* emotion in potential domination (Cawcutt et al., 2020; Manzar et al., 2021). In brief, the *FEAR* emotion increased significantly after the emergence of "*human-to-human transmission of new coronary pneumonia*", a result consistent with the previous research results of Dubey et al. (2020) and Xue et al. (2020).

The intensity of the *ANGRY* emotion expressed by netizens also increased after Event 1. This might be due to the lack of corresponding authoritative information dissemination channels in the early stages, and it was difficult for the public to obtain correct information on related events (Jost et al., 2018). Especially for sudden public health events like the COVID-19 pandemic, the public knew little about it, and they could only learn about the pandemic through online information. However, the uneven information on the Internet will make the public confused about the dynamics of the pandemic (Majeed et al., 2020). In particular, the widespread use of the infodemic on the Internet had further reduced the public's access to reliable information (Medford et al., 2020). Therefore, after the news of the "*human-to-human transmission was confirmed*", netizens will accuse the relevant departments of the untimely release of information and explode with angry expressions.

Fortunately, the active response measures of relevant departments might lead to a significant increase in *POSITIVE ENERGY*. After the COVID-19 pandemic was confirmed, all social sectors extended a helping hand. A steady stream of aid materials began to be collected immediately, and the public began active self-rescue actions (Li et al., 2020). Non-governmental assistance had also been continuously delivered to Wuhan. Such a move has greatly strengthened the public confidence in the fight against the pandemic. Na et al. (2021) pointed out that collectivism can be used as a social defense against the COVID-19 pandemic. China is a typical collectivist country with a high sense of mutual assistance among the nationals. Collectivism means a tightly connected framework in which members are interdependent and expected to care for each other in exchange for unconditional loyalty (Messner & Payson, 2021). Previous studies have shown that collectivism can buffer the high degree of uncertainty and stress caused by the spread of infectious diseases (Kim et al., 2016). Thus, the positive effects of collectivism might be ultimately reflected in the fact that the public showed a high level of positive energy emotional expression for COVID-19.

Further analysis of Event 2 found that, around April 4, 2020, the emotions of *SAD*, *POSITIVE ENERGY* and *FEAR* had increased significantly. MANOVA analysis found that the increase of *SAD* was higher than other emotions. April 4, 2020, was the traditional Chinese Tomb-sweeping Day, a day of sorrow and remembrance. A nationwide ceremony was held on that day to pay tribute to all those who had given their lives in fighting COVID-19, and others who had died of the novel coronavirus (Xinhua News Agency, 2020). On social media platforms, there were lots of condolences for those who

fought against the pandemic and compatriots who died. At the same time, the medium publicity has increased the channels for users to contact the mourning activities (Medford et al., 2020). This kind of nostalgia for the martyrs fighting the pandemic had aroused the resonance of all citizens. Previous studies have pointed out that the *SAD* emotion is an emotional response caused by loss (Raghunathan & Corfman, 2004). Under the COVID-19 pandemic, all citizens might be facing a situation of loss. Whether it was family bereavement or the change of living environment, the emotional expression of *SAD* might be further aggravated. In addition, the increase of *POSITIVE ENERGY* was essentially public confidence in the future of the country and nation. It was the selfless dedication and sacrifice of domestic anti-pandemic heroes that brought about the initial victory in the prevention and control of the COVID-19 pandemic and the safety of public lives and property. This kind of national self-confidence might lead to the increase of *POSITIVE ENERGY*.

In general, different types of events may trigger different emotions among netizens. Previous studies had shown that emotions are situational (Deng et al., 2020). In the face of specific life events, it was easy to induce emotions that were consistent with the situation (Deng et al., 2020). Social events that occurred during the pandemic itself carry certain emotional information, so netizens were more likely to have emotions consistent with the context of the event. However, what we should pay attention to is the emotional expression of *POSITIVE ENERGY*.

Limitation

There exist several shortcomings in this study. First of all, this study chose a semi-automatic method to build the fine-grained sentiment lexicon and supervised ML to analyze emotional intensity. Both of these two methods rely on manual labeling, and subjective bias may exist inevitably. Future research can be conducted in exploration of automated lexicon construction methods and unsupervised ML methods. Secondly, although this study includes the positive energy category, the construction of the positive emotion theory is not sufficient. Subsequent research should fully consider the expression of emotions under the Chinese cultural background, and construct a more comprehensive and complete theory of emotion classification. Finally, our research may face the challenge of ecological validity. Even though the dataset was collected from one of the most popular social media platforms in China, there were still problems such as insufficient population coverage, which make the results hard to generalize to more linguistically diverse environments.

Appendix

Table 7 Data search strategy terms for the original texts

Keyword Terms	
Clue words (COVID-19)	新冠 ("Corona Virus Disease 2019"), 疫情 ("Pandemic"), 新冠疫情 ("COVID-19 Epidemic"), 新冠病毒 ("Coronavirus"), 新冠肺炎 ("Viral Pneumonia"), 肺炎 ("Pneumonia")
AND	
Infectious Noun	人传人 ("Human-To-Human Transmission"), 潜伏期 ("Incubation Period"), 密切接触者 ("Close Contact"), 确诊患者 ("Confirmed Case"), 疑似患者 ("Suspected Case"), 死亡率 ("Mortality Rate"), 治愈率 ("Recovery Rate"), 交叉感染 ("Cross Infection"), 流调 ("Epidemiological Investigation"), 拐点 ("Turning Point")
OR	
Institution Noun	世卫组织 ("World Health Organization, WHO"), 中央应对疫情工作领导小组 ("Leading Group of the CPC Central Committee for Novel Coronavirus Prevention and Control"), 中国疾病预防控制中心 ("Chinese Center for Disease Control and Prevention, China CDC"), 国家卫健委 ("National Health Commission, NHS"), 国务院联防联控机制 ("Joint Prevention and Control Mechanism of the State Council of China")
OR	
Location Noun	中国大陆 ("Mainland China"), 湖北 ("Hubei, a Province in China"), 武汉 ("Wuhan, a City in Hubei Province"), 金银潭医院 ("Wuhan Jinyintan Hospital"), 疫区 ("Affected Area"), 华南海鲜市场 ("Huanan Seafood Market"), 火神山医院 ("Huoshenshan Hospital"), 雷神山医院 ("Leishenshan Hospital"), 定点医院 ("Designated Hospitals"), 方舱医院 ("Temporary Treatment Centers")
OR	
Measure Noun	一级响应 ("First-level Public Health Emergency Response"), 封城 ("Lockdown"), 14天隔离 ("Self-quarantine for 14 days"), 禁止密集聚会 ("Ban Mass Gatherings"), 关闭景点 ("Close Scenic Spots"), 居家办公 ("Work From Home"), 在线教育 ("Online Education"), 口罩 ("Facemask"), 消毒 ("Disinfectant"), 通风 ("Ventilation"), 体温检测 ("Body Temperature Monitoring"), 中医治疗 ("Traditional Chinese Medicine Therapy"), 驰援 ("Assist Wuhan"), 解封 ("Lift The Lockdown"), 疫苗 ("Vaccine"), 医用酒精 ("Medical Alcohol")

the format of keywords is Original Chinese keywords ("English Translation")

Acknowledgements The authors would like to thank all the volunteers from the School of Psychology, Guizhou Normal University.

Author contributions Wenhao Pan analyzed data and wrote up the original draft. Yingying Han designed and supervised the study, reviewed the manuscript. Jinjin Li was involved in the writing of the manuscript. Emily Zhang discussed the results and proofread the manuscript. Bikai He was involved in the data analysis and interpretation. Wenhao Pan and Yingying Han contributed equally and should be recognized as co-first authors. All authors contributed to and have approved the final manuscript.

Funding This work was supported by the National Natural Science Foundation of China (No. 71974058), the Major Research Project of Philosophy and Social Science of Ministry of Education of China (No. 19JZD026), and the Major Research Project of Guizhou Education Reform and Development (No. ZD202007).

Data availability The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Declarations

Ethics approval All procedures performed in studies involving human participants were in accordance with the ethical standards of Guizhou Normal University and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Consent to participate The written informed consent was obtained from all of the participants.

Consent to publish All individual participants have consented to the publication of their data.

Conflict of interest The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

- Aslam, F., Awan, T. M., Syed, J. H., Kashif, A., & Parveen, M. (2020). Sentiments and emotions evoked by news headlines of coronavirus disease (COVID-19) outbreak. *Humanities and Social Sciences Communications*, 7(1), 23. <https://doi.org/10.1057/s41599-020-0523-3>
- An, L., Zhou, W., Ou, M., Li, G., Yu, C., & Wang, X. (2021). Measuring and profiling the topical influence and sentiment contagion of public event stakeholders. *International Journal of Information Management*, 58, 102327. <https://doi.org/10.1016/j.ijinfomgt.2021.102327>
- Bai, X., Chen, F., & Zhan, S. (2014). *A study on sentiment computing and classification of Sina Weibo with Word2vec*. In *Poster at the 2014th IEEE International Congress on Big Data*. <https://doi.org/10.1109/BigData.Congress.2014.59>
- Bandhakavi, A., Wiratunga, N., Massie, S., & P, D. (2021). Emotion-aware polarity lexicons for Twitter sentiment analysis. *Expert Systems*, 38(7), e12332. <https://doi.org/10.1111/exsy.12332>
- Barbalet, J. M. (1998). *Emotion, social theory, and social structure: A macrosociological approach*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511488740>

- Blanken, T. F., Tanis, C. C., Nauta, F. H., Dablander, F., Zijlstra, B. J. H., Bouten, R. R. M., Oostvogel, Q. H., Boersma, M. J., van der Steenhoven, M. V., van Harreveld, F., de Wit, S., & Borsboom, D. (2021). Promoting physical distancing during COVID-19: A systematic approach to compare behavioral interventions. *Scientific Reports*, 11(1), 19463. <https://doi.org/10.1038/s41598-021-98964-z>
- Bollen, J., Mao, H., & Zeng, X. (2011). Twitter mood predicts the stock market. *Journal of Computational Science*, 2(1), 1–8. <https://doi.org/10.1016/j.jocs.2010.12.007>
- Bradley, M. M., & Lang, P. J. (1999). Affective norms for English words (ANEW): Instruction manual and affective ratings (vol. 30, no. 1, pp. 25–36): Technical report C-1, the center for research in psychophysiology, University of Florida.
- Buratto, L. G., Pottage, C. L., Brown, C., Morrison, C. M., & Schaefer, A. (2014). The effects of a distracting n-back task on recognition memory are reduced by negative emotional intensity. *PLoS ONE*, 9(10), e110211. <https://doi.org/10.1371/journal.pone.0110211>
- Cawcutt, K. A., Starlin, R., & Rupp, M. E. (2020). Fighting fear in healthcare workers during the COVID-19 pandemic. *Infection Control & Hospital Epidemiology*, 41(10), 1192–1193. <https://doi.org/10.1017/ice.2020.315>
- Cheng, Q., Li, T. M., Kwok, C., Zhu, T., & Yip, P. S. (2017). Assessing suicide risk and emotional distress in chinese social media: A text mining and machine learning study. *Journal of Medical Internet Research*, 19(7), e243. <https://doi.org/10.2196/jmir.7276>
- China Internet Network Information Center. (2022). *The 48th Statistical Report on China's Internet Development*. CNNIC. Retrieved January 10, 2022, from http://cnnic.cn/guym/xwzx/rdxw/20172017_7086/202202/t20220225_71725.htm
- Crossley, S. A., Kyle, K., & McNamara, D. S. (2017). Sentiment analysis and social cognition engine (SEANCE): An automatic tool for sentiment, social cognition, and social-order analysis. *Behavior Research Methods*, 49(3), 803–821. <https://doi.org/10.3758/s13428-016-0743-z>
- Deng, H., Walter, F., & Guan, Y. (2020). Supervisor-directed emotional labor as upward influence: An emotions-as-social-information perspective. *Journal of Organizational Behavior*, 41(4), 384–402. <https://doi.org/10.1002/job.2424>
- Dong, Z., & Dong, Q. (2003). *HowNet-a hybrid language and knowledge resource*. International Conference on Natural Language Processing and Knowledge Engineering, Proceedings 2003. <https://doi.org/10.1109/NLPKE.2003.1276017>
- Dubey, S., Biswas, P., Ghosh, R., Chatterjee, S., Dubey, M. J., Chatterjee, S., Lahiri, D., & Lavie, C. J. (2020). Psychosocial impact of COVID-19. *Diabetes & Metabolic Syndrome: Clinical Research & Reviews*, 14(5), 779–788. <https://doi.org/10.1016/j.dsx.2020.05.035>
- Duong, D., Ahmad, W. U., Eskin, E., Chang, K., & Li, J. J. (2018). Word and sentence embedding tools to measure semantic similarity of gene ontology terms by their definitions. *Journal of Computational Biology*, 26(1), 38–52. <https://doi.org/10.1089/cmb.2018.0093>
- Ekman, P., & Friesen, W. V. (1971). Constants across cultures in the face and emotion. *Journal of Personality and Social Psychology*, 17(2), 124–129. <https://doi.org/10.1037/h0030377>
- Emodi-Perlman, A., Eli, I., Uziel, N., Smardz, J., Khehra, A., Gilon, E., Wieckiewicz, G., Levin, L., & Wieckiewicz, M. (2021). Public concerns during the COVID-19 lockdown: A multicultural cross-sectional study among internet survey respondents in three countries. *Journal of Clinical Medicine*, 10(8). <https://doi.org/10.3390/jcm10081577>
- Feng, X., Wei, Y., Pan, X., Qiu, L., & Ma, Y. (2020). Academic emotion classification and recognition method for large-scale online learning environment-based on A-CNN and LSTM-ATT deep learning pipeline method. *International Journal of Environmental Research and Public Health*, 17(6), 1941. <https://doi.org/10.3390/ijerph17061941>
- Ferrari, R., Martini, M., Mondini, S., Novello, C., Palomba, D., Scacco, C., Toffolon, M., Valerio, G., Vescovo, G., & Visentin, M. (2009). Pain assessment in non-communicative patients: The Italian version of the Non-Communicative Patient's Pain Assessment Instrument (NOPPAIN). *Aging Clinical and Experimental Research*, 21(4), 298–306. <https://doi.org/10.1007/BF03324919>
- Gao, J., Chen, H., Zhang, X., Guo, J., & Liang, W. (2020). A new feature extraction and recognition method for microexpression based on local non-negative matrix factorization. *Frontiers in Neuroinformatics*, 14, 88. <https://doi.org/10.3389/fnbot.2020.579338>
- Gong, X., Wong, N., & Wang, D. (2018). Are gender differences in emotion culturally universal? Comparison of emotional intensity between Chinese and German samples. *Journal of Cross-Cultural Psychology*, 49(6), 993–1005. <https://doi.org/10.1177/0022022118768434>
- Guo, H., Wang, Y., Zhao, Y., & Liu, H. (2020). Computer-aided design of polyetheretherketone for application to removable pediatric space maintainers. *BMC Oral Health*, 20(1), 201. <https://doi.org/10.1186/s12903-020-01184-6>
- Han, Y., Pan, W., Li, J., Zhang, T., Zhang, Q., & Zhang, E. (2022). Developmental trend of subjective well-being of Weibo users during COVID-19: Online text analysis based on machine learning method. *Frontiers in Psychology*, 12, 779594. <https://doi.org/10.3389/fpsyg.2021.779594>
- Havey, N. F. (2020). Partisan public health: How does political ideology influence support for COVID-19 related misinformation? *Journal of Computational Social Science*, 3(2), 319–342. <https://doi.org/10.1007/s42001-020-00089-2>
- Havnen, A., Anyan, F., Hjemdal, O., Solem, S., Gurigard Riksfjord, M., & Hagen, K. (2020). Resilience moderates negative outcome from stress during the COVID-19 pandemic: A moderated-mediation approach. *International Journal of Environmental Research and Public Health*, 17(18). <https://doi.org/10.3390/ijerph17186461>
- Hoke, T. P., Berger, A. A., Pan, C. C., Jackson, L. A., Winkelman, W. D., High, R., Volpe, K. A., Lin, C. P., & Richter, H. E. (2020). Assessing patients' preferences for gender, age, and experience of their urogynecologic provider. *International Urogynecology Journal*, 31(6), 1203–1208. <https://doi.org/10.1007/s00192-019-04189-0>
- Hutto, D. D., Robertson, I., & Kirchoff, M. D. (2018). A new, better BET: Rescuing and revising basic emotion theory. *Frontiers in Psychology*, 9, 1217. <https://doi.org/10.3389/fpsyg.2018.01217>
- Jost, J. T., Barberá, P., Bonneau, R., Langer, M., Metzger, M., Nagler, J., Sterling, J., & Tucker, J. A. (2018). How social media facilitates political protest: Information, motivation, and social networks. *Political Psychology*, 39(S1), 85–118. <https://doi.org/10.1111/pops.12478>
- Kaufman, B. E. (1999). Emotional arousal as a source of bounded rationality. *Journal of Economic Behavior & Organization*, 38(2), 135–144. [https://doi.org/10.1016/S0167-2681\(99\)00002-5](https://doi.org/10.1016/S0167-2681(99)00002-5)
- Kensinger, E. A., & Corkin, S. (2004). Two routes to emotional memory: Distinct neural processes for valence and arousal. *Proceedings of the National Academy of Sciences of the United States of America*, 101(9), 3310. <https://doi.org/10.1073/pnas.0306408101>
- Keshmiri, S., Shiomi, M., Sumioka, H., Minato, T., & Ishiguro, H. (2020). Gentle versus strong touch classification: Preliminary results, challenges, and potentials. *Sensors*, 20(11). <https://doi.org/10.3390/s20113033>

- Kim, H. S., Sherman, D. K., & Updegraff, J. A. (2016). Fear of Ebola: The influence of collectivism on xenophobic threat responses. *Psychological Science*, 27(7), 935–944. <https://doi.org/10.1177/0956797616642596>
- King, K. M., Molina, B. S. G., & Chassin, L. (2008). A state-trait model of negative life event occurrence in adolescence: Predictors of stability in the occurrence of stressors. *Journal of Clinical Child & Adolescent Psychology*, 37(4), 848–859. <https://doi.org/10.1080/15374410802359643>
- Kirelli, Y., & Arslankaya, S. (2020). Sentiment analysis of shared tweets on global warming on twitter with data mining methods: A case study on Turkish language. *Computational Intelligence and Neuroscience*, 2020, 1904172. <https://doi.org/10.1155/2020/1904172>
- Kosinski, M., Matz, S. C., Gosling, S. D., Popov, V., & Stillwell, D. (2015). Facebook as a research tool for the social sciences: Opportunities, challenges, ethical considerations, and practical guidelines. *American Psychologist*, 70(6), 543–556. <https://doi.org/10.1037/a0039210>
- Kulviwat, S., Bruner, G. C., II., Kumar, A., Nasco, S. A., & Clark, T. (2007). Toward a unified theory of consumer acceptance technology. *Psychology & Marketing*, 24(12), 1059–1084. <https://doi.org/10.1002/mar.20196>
- Kümmel, E., & Kimmerle, J. (2020). The effects of a university's self-presentation and applicants' regulatory focus on emotional, behavioral, and cognitive student engagement. *Sustainability*, 12(23). <https://doi.org/10.3390/su122310045>
- Lee, Y. J., Kamen, C., Margolies, L., & Boehmer, U. (2019). Online health community experiences of sexual minority women with cancer. *Journal of the American Medical Informatics Association*, 26(8–9), 759–766. <https://doi.org/10.1093/jamia/ocz103>
- Leis, A., Ronzano, F., Mayer, M. A., Furlong, L. I., & Sanz, F. (2019). Detecting signs of depression in Tweets in Spanish: Behavioral and linguistic analysis. *Journal of Medical Internet Research*, 21(6), e14199. <https://doi.org/10.2196/14199>
- Li, S., Liu, Z., & Li, Y. (2020). Temporal and spatial evolution of online public sentiment on emergencies. *Information Processing & Management*, 57(2), 102177. <https://doi.org/10.1016/j.ipm.2019.102177>
- Liang, X., Gu, S., Deng, J., Gao, Z., Zhang, Z., & Shen, D. (2015). Investigation of college students' mental health status via semantic analysis of Sina microblog. *Wuhan University Journal of Natural Sciences*, 20(2), 159–164. <https://doi.org/10.1007/s11859-015-1075-z>
- Lin, J., & Yao, Y. (2016). Encoding emotion in Chinese: A database of Chinese emotion words with information of emotion type, intensity, and valence. *Lingua Sinica*, 2(1), 6. <https://doi.org/10.1186/s40655-016-0015-y>
- Luo, K., Deng, Z., Yu, H., & Li, S. (2015). Automatic identification and recognition of sentiment words using an optimization-based model with propagation. *International Journal of Intelligent Systems*, 30(5), 537–549. <https://doi.org/10.1002/int.21707>
- Majeed, M., Irshad, M., Fatima, T., Khan, J., & Hassan, M. M. (2020). Relationship between problematic social media usage and employee depression: A moderated mediation model of mindfulness and fear of COVID-19. *Frontiers in Psychology*, 11, 3368. <https://doi.org/10.3389/fpsyg.2020.557987>
- Manzar, M. D., Albougami, A., Usman, N., & Mamun, M. A. (2021). Suicide among adolescents and youths during the COVID-19 pandemic lockdowns: A press media reports-based exploratory study. *Journal of Child and Adolescent Psychiatric Nursing*, 34(2), 139–146. <https://doi.org/10.1111/jcap.12313>
- Medford, R. J., Saleh, S. N., Sumarsono, A., Perl, T. M., & Lehmann, C. U. (2020). An “Infodemic”: Leveraging high-volume twitter data to understand early public sentiment for the coronavirus disease 2019 outbreak. *Open Forum Infectious Diseases*, 7(7), a258. <https://doi.org/10.1093/ofid/ofaa258>
- Medhat, W., Hassan, A., & Korashy, H. (2014). Sentiment analysis algorithms and applications: A survey. *Ain Shams Engineering Journal*, 5(4), 1093–1113. <https://doi.org/10.1016/j.asej.2014.04.011>
- Messner, W., & Payson, S. E. (2021). Contextual factors and the COVID-19 outbreak rate across U.S. counties in its initial phase. *Health Science Reports*, 4(1), e242. <https://doi.org/10.1002/hsr2.242>
- Min, H., Peng, Y., Shoss, M., & Yang, B. (2021). Using machine learning to investigate the public's emotional responses to work from home during the COVID-19 pandemic. *Journal of Applied Psychology*, 106(2), 214–229. <https://doi.org/10.1037/apl0000886>
- Moon, E., Yang, M., Seon, Q., & Linnaranta, O. (2021). Relevance of objective measures in psychiatric disorders—rest-activity rhythm and psychophysiological measures. *Current Psychiatry Reports*, 23(12), 85. <https://doi.org/10.1007/s11920-021-01291-3>
- Mozetič, I., Grčar, M., & Smailović, J. (2016). Multilingual Twitter sentiment classification: The role of human annotators. *PLoS ONE*, 11(5), e155036. <https://doi.org/10.1371/journal.pone.0155036>
- Muraoka, K., Sada, Y., Miyazaki, D., Chaikittisilp, W., & Okubo, T. (2019). Linking synthesis and structure descriptors from a large collection of synthetic records of zeolite materials. *Nature Communications*, 10(1), 4459. <https://doi.org/10.1038/s41467-019-12394-0>
- Na, J., Kim, N., Suk, H. W., Choi, E., Choi, J. A., Kim, J. H., Kim, S., & Choi, I. (2021). Individualism-collectivism during the COVID-19 pandemic: A field study testing the pathogen stress hypothesis of individualism-collectivism in Korea. *Personality and Individual Differences*, 183, 111127. <https://doi.org/10.1016/j.paid.2021.111127>
- Naslund, J. A., Bondre, A., Torous, J., & Aschbrenner, K. A. (2020). Social media and mental health: Benefits, risks, and opportunities for research and practice. *Journal of Technology in Behavioral Science*, 5(3), 245–257. <https://doi.org/10.1007/s41347-020-00134-x>
- Nicomedes, C. J. C., & Avila, R. M. A. (2020). An analysis on the panic during COVID-19 pandemic through an online form. *Journal of Affective Disorders*, 276, 14–22. <https://doi.org/10.1016/j.jad.2020.06.046>
- Novembre, G., Zanon, M., Morrison, I., & Ambron, E. (2019). Bodily sensations in social scenarios: Where in the body? *PLoS ONE*, 14(6), e206270. <https://doi.org/10.1371/journal.pone.0206270>
- Ostic, D., Qalati, S. A., Barbosa, B., Shah, S. M. M., Galvan Vela, E., Herzallah, A. M., & Liu, F. (2021). Effects of social media use on psychological well-being: A mediated model. *Frontiers in Psychology*, 12, 678766. <https://doi.org/10.3389/fpsyg.2021.678766>
- Pennebaker, J. W., Boyd, R. L., Jordan, K., & Blackburn, K. (2015). *The development and psychometric properties of LIWC2015*. University of Texas at Austin. Retrieved February 10, 2022, from <http://hdl.handle.net/2152/31333>
- Power, M. J. (2006). The structure of emotion: An empirical comparison of six models. *Cognition and Emotion*, 20(5), 694–713. <https://doi.org/10.1080/02699930500367925>
- Pröllochs, N., Feuerriegel, S., & Neumann, D. (2018). Statistical inferences for polarity identification in natural language. *PLoS ONE*, 13(12), e209323. <https://doi.org/10.1371/journal.pone.0209323>
- Raghunathan, R., & Corfman, K. P. (2004). Sadness as pleasure-seeking prime and anxiety as attentiveness prime: The “Different Affect-Different Effect” (DADE) model. *Motivation*

- and Emotion, 28(1), 23–41. <https://doi.org/10.1023/B:MOEM.0000027276.32709.30>
- Raghupathi, V., Ren, J., & Raghupathi, W. (2020). Studying public perception about vaccination: A sentiment analysis of Tweets. *International Journal of Environmental Research and Public Health*, 17(10). <https://doi.org/10.3390/ijerph17103464>
- Ranieri, J., Guerra, F., Martelli, A., Fanelli, V., & Di Giacomo, D. (2021). Impact of cybersex and intensive Internet use on the well-being of Generation Z: An analysis based on the EPOCH model. *Journal of Technology in Behavioral Science*, 6(3), 501–506. <https://doi.org/10.1007/s41347-021-00197-4>
- Ready, R. E., Åkerstedt, A. M., & Mroczek, D. K. (2012). Emotional complexity and emotional well-being in older adults: Risks of high neuroticism. *Aging & Mental Health*, 16(1), 17–26. <https://doi.org/10.1080/13607863.2011.602961>
- Reisenzein, R. (1994). Pleasure-arousal theory and the intensity of emotions. *Journal of Personality and Social Psychology*, 67(3), 525–539. <https://doi.org/10.1037/0022-3514.67.3.525>
- Rodríguez, F. M., & Garza, S. E. (2019). Predicting emotional intensity in social networks. *Journal of Intelligent & Fuzzy Systems*, 36, 4709–4719. <https://doi.org/10.3233/JIFS-179020>
- Roh, T., Jeong, Y., Jang, H., & Yoon, B. (2019). Technology opportunity discovery by structuring user needs based on natural language processing and machine learning. *PLoS ONE*, 14(10), e223404. <https://doi.org/10.1371/journal.pone.0223404>
- Rubin, G. J., & Wessely, S. (2020). The psychological effects of quarantining a city. *BMJ*, 368, m313. <https://doi.org/10.1136/bmj.m313>
- Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, 39(6), 1161–1178. <https://doi.org/10.1037/h0077714>
- Sailunaz, K., & Alhajj, R. (2019). Emotion and sentiment analysis from Twitter text. *Journal of Computational Science*, 36, 101003. <https://doi.org/10.1016/j.jocs.2019.05.009>
- Salathé, M., & Khandelwal, S. (2011). Assessing vaccination sentiments with online social media: Implications for infectious disease dynamics and control. *PLoS Computational Biology*, 7(10), e1002199. <https://doi.org/10.1371/journal.pcbi.1002199>
- Salgado, S., & Kingo, O. S. (2019). How is physiological arousal related to self-reported measures of emotional intensity and valence of events and their autobiographical memories? *Consciousness and Cognition*, 75, 102811. <https://doi.org/10.1016/j.concog.2019.102811>
- Sasaki, Y., Rea, B., & Ananiadou, S. (2009). Clinical text classification under the open and closed topic assumptions. *International Journal of Data Mining and Bioinformatics*, 3(3), 299–313. <https://doi.org/10.1504/IJDMB.2009.026703>
- Shaikh, S., Cho, K., Strzalkowski, T., Feldman, L., Lien, J., Liu, T., & Broadwell, G. A. (2016). ANEW+: automatic expansion and validation of affective norms of words lexicons in multiple languages. Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16), Portorož, Slovenia. Retrieved March 1, 2022, from <https://aclanthology.org/L16-1180>
- Shi, Y., Li, P., Yu, X., Wang, H., & Niu, L. (2018). Evaluating doctor performance: Ordinal regression-based approach. *Journal of Medical Internet Research*, 20(7), e240. <https://doi.org/10.2196/jmir.9300>
- Sina Weibo Data Center. (2021). *Weibo 2020 User Development Report*. Sina Weibo Reports. Retrieved November 10, 2021, from <https://data.weibo.com/report/reportDetail?id=456>
- Sønderskov, K. M., Dinesen, P. T., Santini, Z. I., & Østergaard, S. D. (2020). The depressive state of Denmark during the COVID-19 pandemic. *Acta Neuropsychiatrica*, 32(4), 226–228. <https://doi.org/10.1017/neu.2020.15>
- Sonnemans, J., & Frijda, N. H. (1994). The structure of subjective emotional intensity. *Cognition and Emotion*, 8(4), 329–350. <https://doi.org/10.1080/02699939408408945>
- Sonnemans, J., & Frijda, N. H. (1995). The determinants of subjective emotional intensity. *Cognition and Emotion*, 9(5), 483–506. <https://doi.org/10.1080/02699939508408977>
- Stone, P. J., Dunphy, D. C., & Smith, M. S. (1966). *The general inquirer: A computer approach to content analysis*. M.I.T. Press.
- Tavares, P., Barnard, P. J., & Lawrence, A. D. (2011). Emotional complexity and the neural representation of emotion in motion. *Social Cognitive and Affective Neuroscience*, 6(1), 98–108. <https://doi.org/10.1093/scan/nsq021>
- The State Council Information Office of China. (2020). *Fighting COVID-19: China in Action*. White Paper. Retrieved November 13, 2021, from http://english.scio.gov.cn/node_8018767.html
- Twenge, J. M., & Joiner, T. E. (2020). U.S. census bureau-assessed prevalence of anxiety and depressive symptoms in 2019 and during the 2020 COVID-19 pandemic. *Depression and Anxiety*, 37(10), 954–956. <https://doi.org/10.1002/da.23077>
- Wang, C., Pan, R., Wan, X., Tan, Y., Xu, L., Ho, C. S., & Ho, R. C. (2020). Immediate psychological responses and associated factors during the initial stage of the 2019 coronavirus disease (COVID-19) pandemic among the general population in China. *International Journal of Environmental Research and Public Health*, 17(5). <https://doi.org/10.3390/ijerph17051729>
- Warriner, A. B., Kuperman, V., & Brysbaert, M. (2013). Norms of valence, arousal, and dominance for 13,915 English lemmas. *Behavior Research Methods*, 45(4), 1191–1207. <https://doi.org/10.3758/s13428-012-0314-x>
- World Health Organization. (2020). *Report of the WHO-China Joint Mission on Coronavirus Disease 2019 (COVID-19)*. WHO Headquarters. Retrieved November 11, 2021, from [https://www.who.int/publications/i/item/report-of-the-who-china-joint-mission-on-coronavirus-disease-2019-\(covid-19\)](https://www.who.int/publications/i/item/report-of-the-who-china-joint-mission-on-coronavirus-disease-2019-(covid-19))
- Xinhua News Agency. (2020). *Chinese national flag flies at half-mast to mourn for people died in COVID-19 fight*. Xinhuanet. Retrieved December 3, 2021, from http://www.xinhuanet.com/english/2020-04/04/c_138945730.htm
- Xiong, Y., Wang, Z., Jiang, D., Wang, X., Chen, Q., Xu, H., Yan, J., & Tang, B. (2019). A fine-grained Chinese word segmentation and part-of-speech tagging corpus for clinical text. *BMC Medical Informatics and Decision Making*, 19(Suppl 2), 66. <https://doi.org/10.1186/s12911-019-0770-7>
- Xu, L., Li, L., Jiang, Z., Sun, Z., Wen, X., Shi, J., Sun, R., & Qian, X. (2021). A novel emotion lexicon for Chinese emotional expression analysis on Weibo: Using grounded theory and semi-automatic methods. *IEEE Access*, 9, 92757–92768. <https://doi.org/10.1109/ACCESS.2020.3009292>
- Xu, L., Lin, H., Pan, Y., Ren, H., & Chen, J. (2008). Constructing the affective lexicon ontology. *Journal of the China Society for Scientific and Technical Information*, 27(2), 180–185. <https://doi.org/10.3969/j.issn.1000-0135.2008.02.004>
- Xue, J., Chen, J., Hu, R., Chen, C., Zheng, C., Su, Y., & Zhu, T. (2020). Twitter discussions and emotions about the COVID-19 pandemic: Machine learning approach. *Journal of Medical Internet Research*, 22(11), e20550. <https://doi.org/10.2196/20550>
- Yang, T., Xie, J., Li, G., Mou, N., Li, Z., Tian, C., & Zhao, J. (2019). Social media big data mining and spatio-temporal analysis on public emotions for disaster mitigation. *ISPRS International Journal of Geo-Information*, 8(1). <https://doi.org/10.3390/ijgi8010029>
- Ye, Y., Long, T., Liu, C., & Xu, D. (2020). The effect of emotion on prosocial tendency: The moderating effect of pandemic severity under the outbreak of COVID-19. *Frontiers in Psychology*, 11, 588701. <https://doi.org/10.3389/fpsyg.2020.588701>

- Ye, Y., Xu, Y., Zhu, Y., Liang, J., Lan, T., & Yu, M. (2016). The characteristics of moral emotions of Chinese netizens towards an anthropogenic hazard: A sentiment analysis on Weibo. *Acta Psychologica Sinica*, 48(3), 290–304. <https://doi.org/10.3724/SP.J.1041.2016.00290>
- You, J., Expert, P., & Costelloe, C. (2021). Using text mining to track outbreak trends in global surveillance of emerging diseases: ProMED-mail. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 184(4), 1245–1259. <https://doi.org/10.1111/rssa.12721>
- Zeng, R., & Zhu, D. (2019). A model and simulation of the emotional contagion of netizens in the process of rumor refutation. *Scientific Reports*, 9(1), 14164. <https://doi.org/10.1038/s41598-019-50770-4>
- Zhang, Y., & Yu, F. (2018). Which socio-economic indicators influence collective morality? Big data analysis on online Chinese social media. *Emerging Markets Finance and Trade*, 54(4), 792–800. <https://doi.org/10.1080/1540496X.2017.1321984>
- Zhao, D., Ding, R., Zhang, H., Zhang, N., Hu, L., & Luo, W. (2021). Individualized prediction of females' empathic concern from intrinsic brain activity within general network of state empathy. *Cognitive, Affective, & Behavioral Neuroscience*. <https://doi.org/10.3758/s13415-021-00964-z>
- Zhu, J., & Xu, C. (2021). Sina microblog sentiment in Beijing city parks as measure of demand for urban green space during the COVID-19. *Urban Forestry & Urban Greening*, 58, 126913. <https://doi.org/10.1016/j.ufug.2020.126913>
- Zhu, Y., Yan, E., & Wang, F. (2017). Semantic relatedness and similarity of biomedical terms: Examining the effects of recency, size, and section of biomedical publications on the performance of word2vec. *BMC Medical Informatics and Decision Making*, 17(1), 95. <https://doi.org/10.1186/s12911-017-0498-1>

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

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.