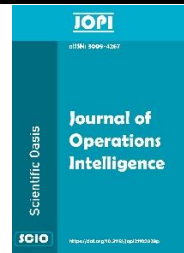




SCIENTIFIC OASIS

Journal of Operations Intelligence

Journal homepage: www.jopi-journal.org
eISSN: 3009-4267



Chinese Sentiment Analyses Research: A Systematic Review and Bibliometric Analysis

Shahd Al-Majali¹, Ibrahim Alshakhatreh^{1,*}

¹ Department of Business Administration, College of Management, National Yunlin University of Science and Technology, 123, University Road, Section 3, Douliu, Yunlin County, Taiwan

ARTICLE INFO

Article history:

Received 29 October 2024
Received in revised form 29 January 2025
Accepted 22 March 2025
Available online 31 March 2025

Keywords:

Opinion mining; Opinion analysis; Chinese language; Taiwanese language.

ABSTRACT

This study aims to shed light on topics related to sentiment analysis in the Chinese language, due to the scarcity of studies in this area and its growing importance. We used a systematic, bibliometric approach for its effectiveness in analyzing the current state of sentiment analysis research in Chinese. The study identifies three main stages based on the analysis of previous research in this field. In the first stage, we examine the challenges faced by researchers in analyzing sentiment in Chinese, including the lack of access to a comprehensive lexicon. In the second stage, we discuss the motives underlying the growing importance and potential impact of sentiment analysis in Chinese. Finally, we make recommendations to guide future research in this field to improve the accuracy and reliability of sentiment analysis tools for Chinese.

1. Introduction

The field of sentiment analysis has gained an important place for research in recent years [1-4], by virtue of its effective role in understanding people's opinions and feelings towards products [5-7], services [8], and events [5, 9]. Thanks to the availability of social media platforms, especially through e-commerce sites, companies and institutions are interested in understanding and analyzing these feelings to gain a clearer understanding of consumer behavior and predict the customer's list of preferences [6, 8]. Sentiment analysis is increasingly available in English with interest in Chinese sentiment analysis [10], and sentiment analysis in Chinese poses unique challenges due to its complex structure [4, 8] being a language that contains a huge amount of idioms and metaphors [6]. This shows that Chinese sentiment analysis is getting exciting and growing attention for its development, especially when integrating the latest technologies with advanced methods [10], such as artificial intelligence and machine learning [7, 11]. Because of their effective ability to analyze the huge amounts of available data through social media platforms in various sectors [12], they can gain

* Corresponding author.

E-mail address: ibshakhatreh@gmail.com

<https://doi.org/10.31181/jopi31202536>

© The Author(s) 2025 | [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/)

valuable insights into the sentiments of public opinion [8]. Moreover, these new technologies help to develop new algorithms and help to identify them, thus facilitating the accurate extraction of emotions [13]. Despite the growing interest in developing methods and capacity for sentiment analysis in Chinese, researchers are still facing significant challenges [14]. One of them is the scarcity of high-quality research and the scarcity of disaggregated data for sentiment analysis in Chinese [1, 2, 10]. In addition, the characteristics that distinguish Chinese from other languages, such as the diversity of Chinese dialects and the complexity of Chinese expressions and metaphors, make its analysis difficult for researchers [13, 15]. Therefore, researchers must work on developing advanced techniques to address these challenges and improve the accuracy of Chinese sentiment analysis [16]. To this end, this paper presents a comprehensive overview of an article review on Chinese sentiment analysis and discusses the challenges and motivations of Chinese sentiment analysis with some recommendations for future research.

The paper begins with a discussion of the challenges associated with analyzing Chinese feelings, such as the problems that the researcher may face due to the nature of the data and the characteristics that characterize the Chinese language. Certainly, it discusses the difficulty of accurately defining feelings in relation to Chinese expressions and metaphors. Based on the experiences of previous researchers. Secondly, discussing the importance of Chinese sentiment analysis mentioned some of its potential benefits that may affect various sectors. Thirdly, I will provide some advice based on what was recommended by previous researchers.

2. Systematic Review

This literature is based on a comprehensive and reproducible methodology using a systematic review protocol [9]. This approach included identifying all relevant topics in the same field and defining a strategy, inclusion criteria, and data extraction methods. By doing so, this methodology minimized the risk of excluding important studies or including irrelevant ones, thus enhancing the credibility and validity of the findings. The systematic approach also helped minimize bias and subjective interpretations, resulting in more accurate and reliable results. Furthermore, this approach saved time and increased efficiency by streamlining the review process. In summary, the systematic review protocol used in this study ensured a clear, systematic, and comprehensive literature review, enhancing the ability to make informed decisions based on credible and accurate results [11, 17, 18].

2.1 Information Source

A source of information is an essential component of research, providing valuable insights and data to support a thesis or hypothesis [17]. This not only applies to academic studies but also to a wider area of research across different disciplines. After consultation on the field of study, this study included four digital sources: IEEE, Wiley Online Library, Science Direct, and Web of Science, which are scientifically recognized sources for their comprehensiveness and reliability in the academic community. IEEE is a professional association and publisher with a wealth of resources that assist computer scientists on related technology and engineering research. Wiley Online Library is a digital library that gives researchers access to thousands of academic journals, books, and reference works across a wide range of fields of study [17, 18]. Science Direct contains a large collection of scientific and technical articles, books, and papers from thousands of reputable journals and book series. Web of Science is a citation index that allows users to explore research connections and track citation trends across various disciplines, providing valuable insights into the latest developments in their field. Researchers can draw on these sources to find reliable information to support their research and advance their understanding of the topic of their choice [18].

2.2 Search Strategy

To ensure a comprehensive search in the literature, some procedures were carried out, which are the identification stage, the examination stage, the eligibility stage, and the inclusion stage, as seen in Figure 1 SLR Protocol, starting with the definition stage in August 2022 after identifying the main points of the manuscript, and the second stage in September 2022 to include more current and updated sources are included. Using the electronic search box available on the scientific site referred to above, and to organize the search process on the scientific references to be used, (AND, OR) queries were used among the keywords that help to sort more specifically [17, 18]. To conduct the search, there are two sets of keywords or queries were used, as shown in Figure 1: ("Sentiment Analysis" OR "Opinion Mining" OR "Emotional Analysis" OR "Opinion Analysis") AND ("Chinese language" OR "Taiwanese language"), which covers the period 2009-2022. It is important to note that only articles written in English were included in the search.

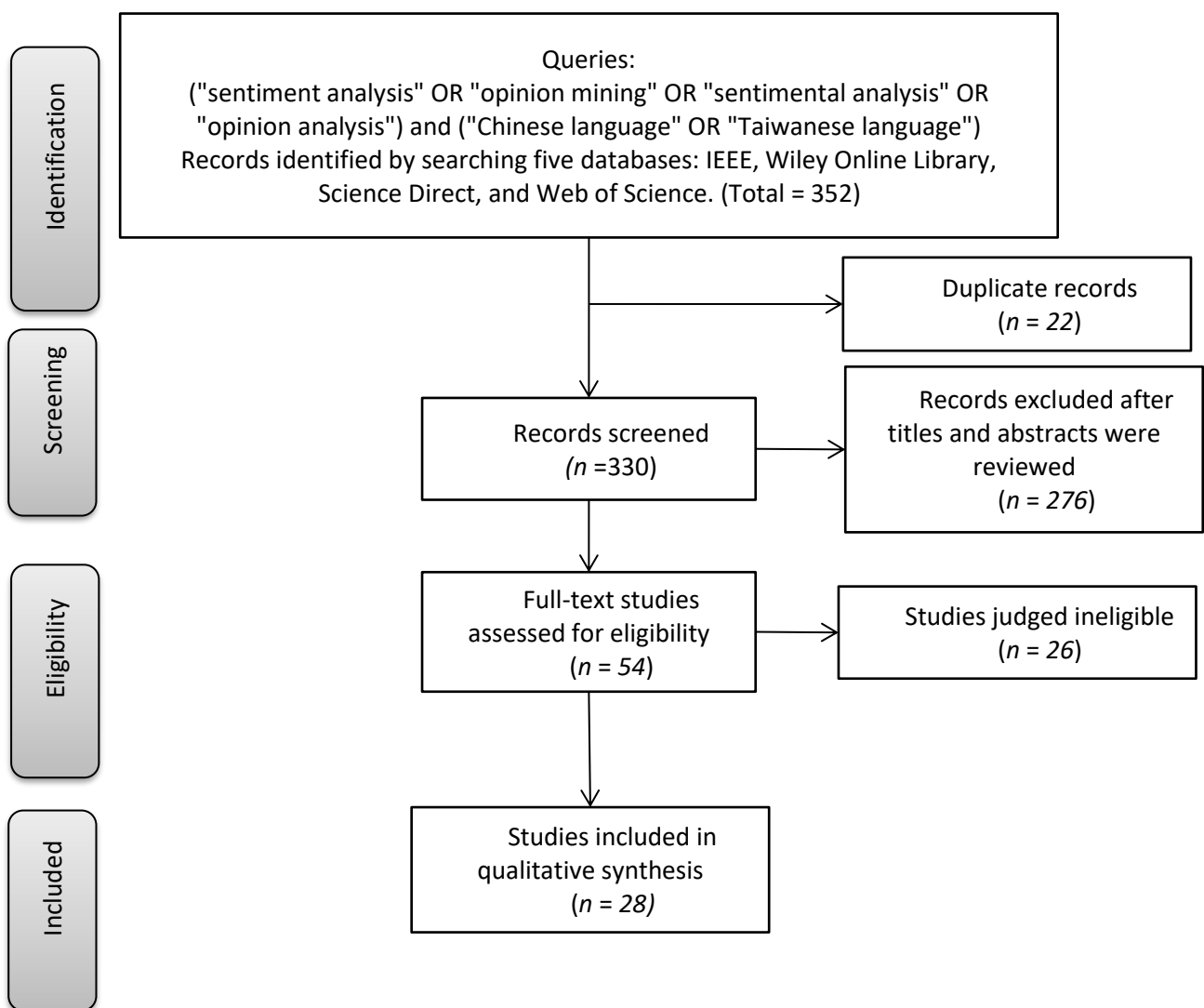


Fig. 1. SLR Protocol

2.3 Study Selection

As seen in Figure 1, The research process comprised four distinct sub-processes, after identifying the main points of the manuscript as the first process, then namely article collection, title and abstract scanning, and full-text reading [17, 18]. Initially, a total of (n = 352) articles were collected from the selected databases, and after scanning, (n = 22) duplicate articles were identified. The subsequent step involved scanning the abstracts and titles of the collected articles to determine their relevance. A thorough investigation of the titles and abstracts was conducted to ensure that eligible articles fully met the inclusion criteria established before. A full-text reading was performed in the final sub-process, and any study that did not meet the criteria was excluded. A total of (n = 276) unrelated papers were identified during this step. Useful and valuable information components (i.e., data extraction) were extracted from the final set of articles to shape the final form of the review, as presented in Table 1.

Table 1

Data extraction elements

Data Item	Description
Title	The Main Topic or Focus of The Research
Year	The Year in Which the Research Article Was Published
Aim	The Purpose of The Study
Motivations	Explains Why the Researchers Undertook the Study
Challenges	Difficulties the Researchers Faced During the Study
Recommendation - Future Work	Highlighting Areas Where Further Investigation Is Needed or Suggesting Improvements to The Methodology
Journal Name	Refers to the Academic Journal in Which the Research Article Was Published
Data Source	The Source of The Data Used in The Study
Method	Section Explains the Research Methodology and Procedures Used in The Study

As shown in Table 1, these are the essential elements in the process of analyzing the data extracted from the articles identified in the previous stage. We start with the article title that can be used, followed by the year of publication. Which one can be used? Third, we deduce the study's objective to investigate a particular research question or hypothesis. Fourth, identify the motivations behind the study to provide insight into why the researchers decided to conduct the study. Fifth. The challenges the researchers faced during the study provide context for the limitations of the research. Sixth, recommendations and areas needing further investigation or suggestions for methodological improvement are highlighted. Seventh: The name of the journal in which the article was published so readers can identify the original source of the research. Data Source Eight About the source, the researchers obtained the data used in the study. Finally, extract the research methodology and procedures involved in the study.

2.4 Inclusion and Exclusion Criteria

Inclusion and exclusion criteria are important factors in sentiment analysis in Chinese. In the context of this research, the main inclusion and exclusion criteria can be computed into four steps: first, we excluded studies that included languages other than Chinese, and second, we excluded studies that used a different approach other than sentiment analysis. Thirdly, the focus was only on studies for Chinese Sentiment analysis, emphasizing articles written in English from 2009 to 2022. Fourthly, the required scientific research was extracted, and a table was created, which includes the most attractive elements for deciphering the primary motivations of the researchers; thus this approach ensured the collection of the latest research and aims to ensure a high standard of accuracy and efficiency [17, 18].

3. Bibliometric analysis

Bibliometric analysis is a powerful tool with significant implications for research and academia. It can help identify emerging research directions and collaborations between researchers, and countries in specific fields, providing valuable insights into research performance, productivity, and collaboration networks. This allows researchers and institutions to better understand the research landscape, improve search results, and foster new research collaborations. To perform bibliometric analysis, a set of specific steps must be followed. First, identifying the database in this study, the Scopus database was used to extract a file containing information about scientific articles. Secondly, the information that needs to be extracted is selected, and the data is exported using the BibTeX methodology. This results in a file that contains the database of global articles specified for the study in Scopus format. Thirdly, the bibliometric, library(bibliometric), and bibliophagy packages are installed in R Studio, and the "Run" icon is clicked. Finally, the web page for bibliometric analysis based on the bibliometrics methodology is automatically opened. The extracted database is uploaded as a data source, and the analysis of scientific articles downloaded from the database is displayed. Bibliometric analysis is an invaluable tool for researchers and institutions seeking to enhance the performance and impact of their research by gaining deeper insights into research trends, productivity, collaborations, and impact. According to Figure 2 (Annual Scientific Production) after the bibliometric analysis, which appears based on the citations and the selected years (2009-2022). It is the growth and interest of researchers in sentiment analysis in the Chinese language. Also, this analysis can show us the relationship between keywords, countries, and researchers interested in the field of study, allowing the opportunity to gain a deeper understanding of these relationships.as shown in Figure 3.

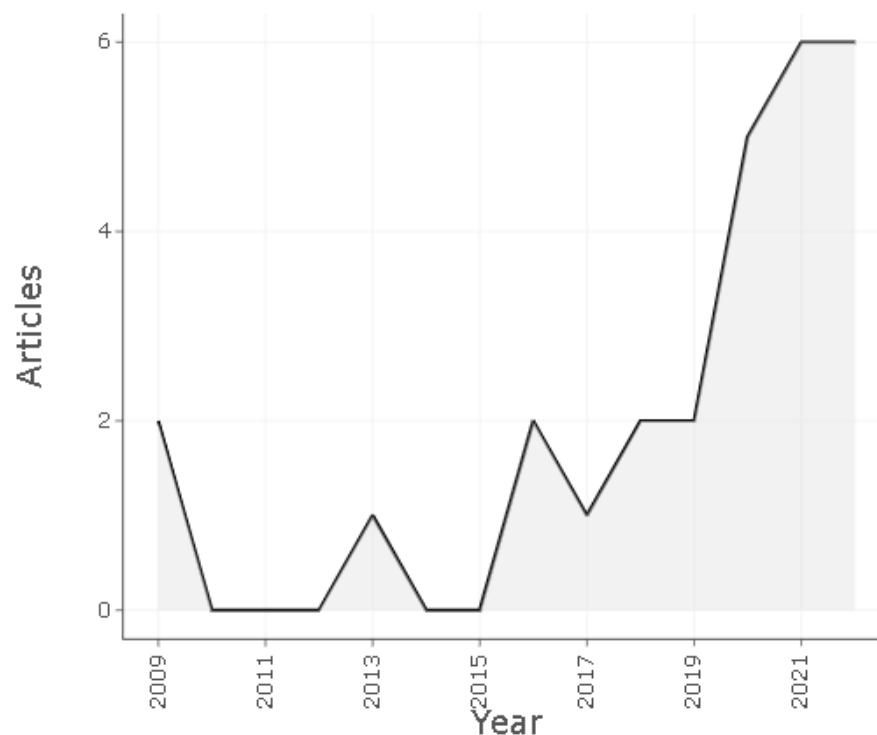


Fig. 2. Annual Scientific Production

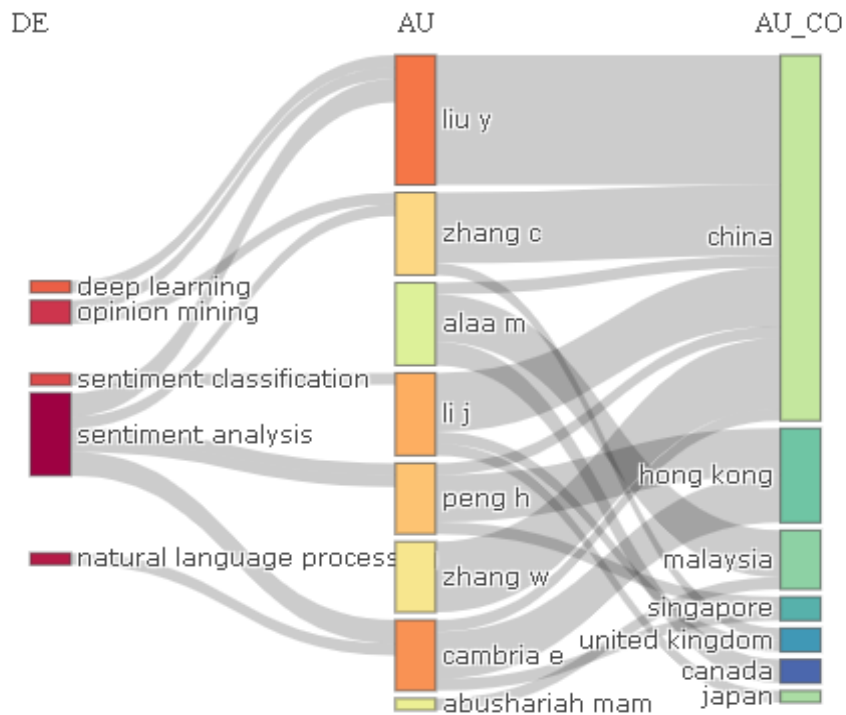


Fig. 3. presents an overview of the relationship between keywords, countries, and researchers interested in the field of X

In addition, this type of analysis has the ability to provide research information on a topic by getting to know the journal, country, or researchers who can collaborate to come up with the best study result. As shown in Figure 4 and Figure 5 by providing us with a list of countries and journals interested in publishing the topic, based on the database provided by the program that helps determine the field in which the researcher wants to include his research.

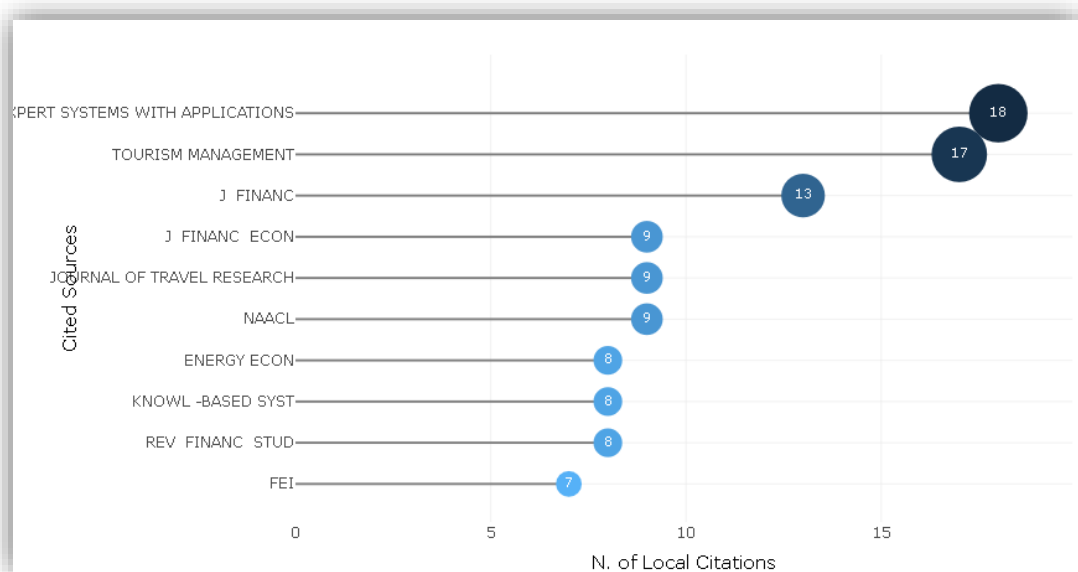


Fig. 4. Name of most journal cited related to the same topic

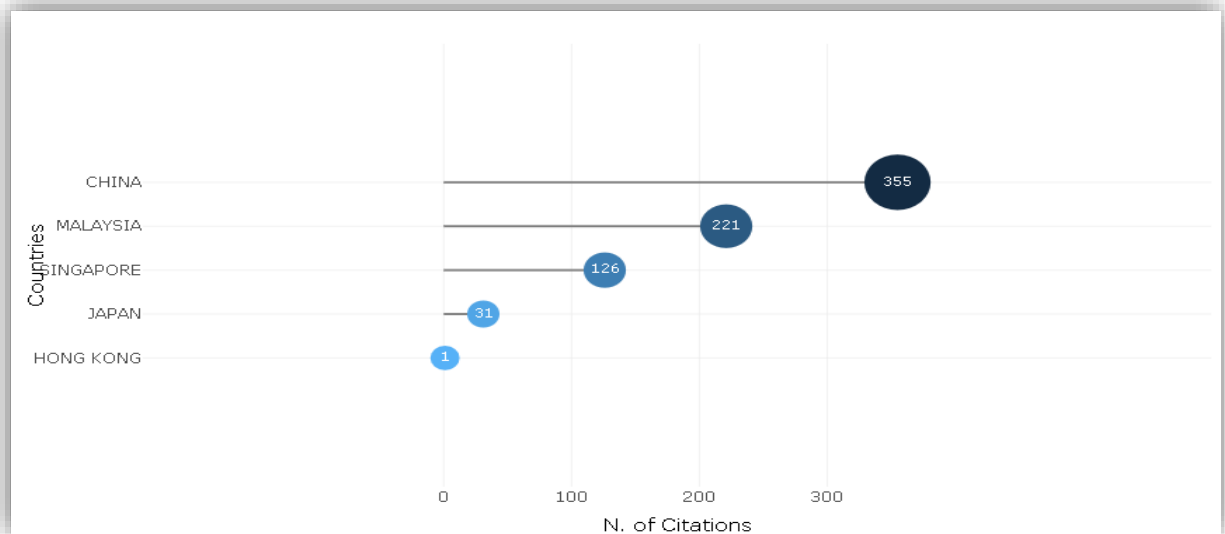


Fig. 5. Most Cited Countries

Furthermore, applying the bibliometric methodology has facilitated the comprehensive identification of commonly repeated keywords and the strength of their association in research studies. This approach helps researchers choose relevant keywords and enhance the overall value of their study. The use of bibliometric analysis is highlighted in Figure 6, which depicts the keywords used in the international articles identified in this study. It should be noted that the process of bibliometric analysis is notorious for providing information about the network of collaborations between authors and countries, as can be seen in Figure 7.

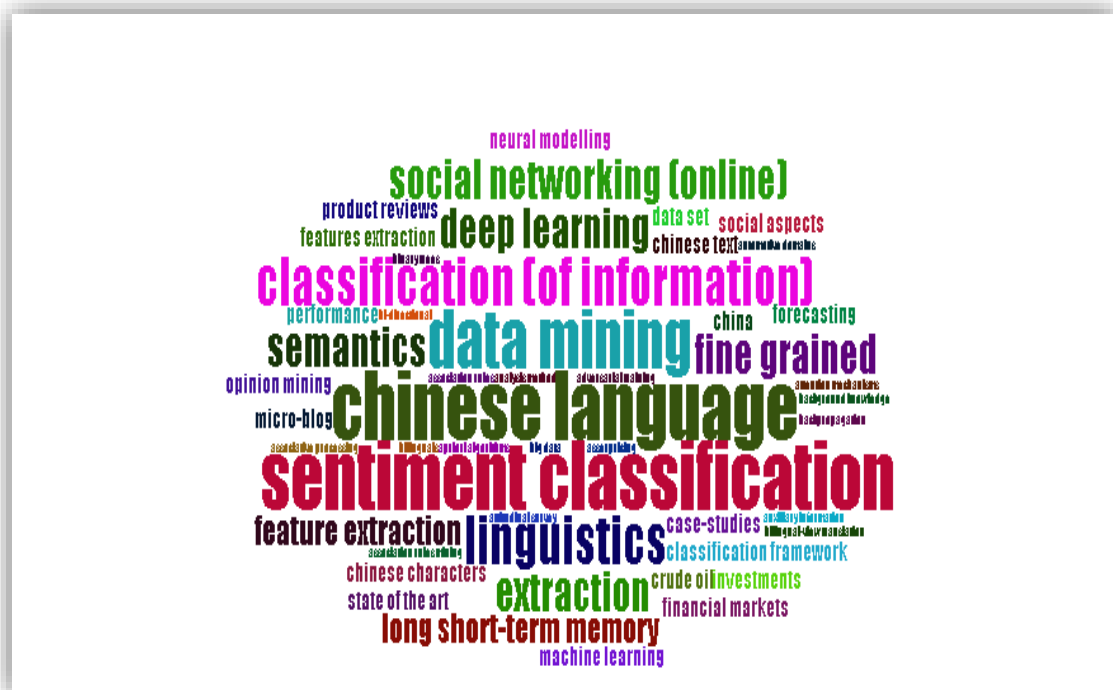


Fig. 6. Most Frequent Words (The larger the keyword, the more frequently it is used in studies)

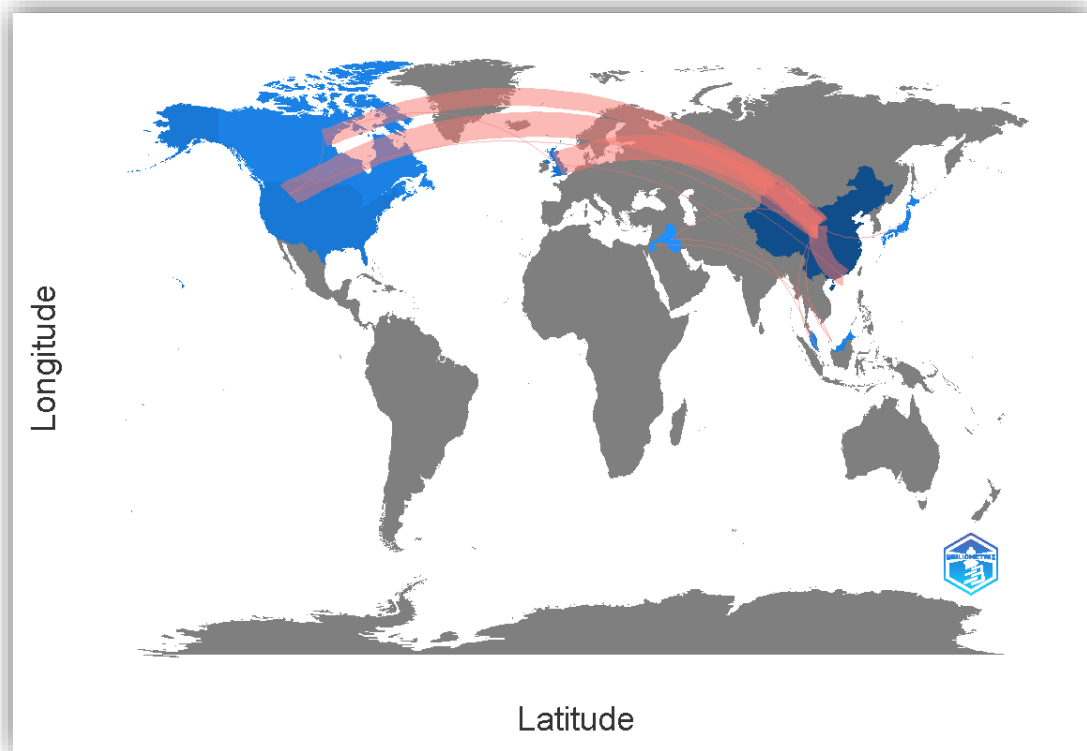


Fig. 7. Collaboration by using World Map (the red line connects the countries that cooperate together)

4. Challenges

sentiment analysis is a difficult analysis due to the difficulty of extracting emotions, especially hidden emotions, and during our study on Chinese sentiment analysis, we will discuss four main headlines of Chinese sentiment analysis challenges: 1- The nature of the Chinese language, 2- The role of social media platforms, 3- challenge of research related to emotions. 4- the technical techniques challenges

4.1 Language

Our study includes that there are many challenges that researchers faced in studying g Chinese sentiment analysis, and language is one of them, as it is the main bridge for expressing the feelings and desires of individuals and groups, which consists of backward characteristics and rules that distinguish one language from another. While identifying the academic literature, it was found that language has two main sub-challenges, which include, (1) the rules of language and its (2) characteristics.

4.1.1 Language Rules

Each language has rules and grammar that help control sounds, words, and dialogue, making learning and understanding its information easier. Despite that, language in SA research from the rule's perspective has its own pitfalls. In the context of our study, the researchers pointed out the difficulty of performing traditional sentiment analysis in Chinese. First, regarding semantic information and grammatical features, the researchers emphasized that Chinese sentiment analysis is difficult in terms of semantic meaning, as the Chinese language features each word with special semantic information [19]. Sometimes, it is a positive word, but it is an expression of negative feelings

in semantic terms [20]. Second, the unique sentence grammar. In that regard, it is fair to say that Chinese is one of the languages that characterizes writing with no spaces between words, and it consists of words having numerous syllables "Polysyllabic". These features turn out to be a challenge for the development of sentiment analysis [21]. Third, the variety of syntactic dependency is the relationship between two words in a sentence, where one word is the ruler and the other depends on the relationship. The fact that the language does not contain spaces, as mentioned previously. It leads to more grammatical meanings, making it difficult to know the grammatical meaning and determine the nature of feelings [22].

4.1.2 Language Characters

Chinese language characters use different linguistic strategies, which has been identified as challenging when conducting SA for the Chinese language. There are several issues in that regard, and firstly, it has few characters and a large number of words, which are used to form groups to form several other words [23]. Second, the combination of characters is challenging given the fact that a character can have more than one meaning. The words or expressions are formed from many combinations of these characters, and this further complicates sentiment analysis. In addition, Chinese as an analytic language is a language that conveys the relationships between words in sentences primarily by means of auxiliary words as opposed to using inflections [19]. Third, Chinese is the language of the diagram because its characters correspond to the pictograms [15].

4.2 Social Media Platform

The world has become like a small village, and we can access and learn about all the events, news and experiences of others without having to get to know them personally, with different (races, genders, places, religions, etc.). This is due to social media platforms, the second challenge researchers face in our study. Since it is considered the fastest way to express feelings and opinions from different points of view in the world through various platforms such as (Facebook, Twitter, Instagram ... and others), but it faces the problem of content in terms of its (1) nature and the difficulty of (2) processing to obtain useful information quickly and sometimes the abundance of information that is not useful or unrealistic.

4.3 Nature of Online Content

Social media platforms have become integral to our daily lives, providing us with endless information and opportunities to connect with others. However, the challenge of Chinese sentiment analysis on social media platforms is a complex issue that involves various aspects of online content, including bias, massive information, structure, translation system, emotional tendencies, and data retrieval. First, bias is an inherent problem with online content because individuals or groups largely create it with their own personal opinions, beliefs, and values. This can lead to the creation of biased content, which may make it difficult to identify the real feeling behind the content, because it may not be a true reflection of the public's opinion and one of the policies that have been used to promote or hide feelings across social media platforms [23]. Second, the online content on social media platforms generated every second includes a huge amount of content, making it almost impossible to manually analyze every single post, tweet, or comment. [13]. This has made it necessary to have powerful automated sentiment analysis tools to handle large amounts of data. Third, online content structure may be inconsistent and disorganized [21], and posts can range from simple text messages to long texts and multimedia content such as photos or videos [24]. This can make it difficult to apply sentiment analysis techniques and create a standardized approach for effective content analysis [21]. Fourth, the challenge of analyzing Chinese sentiment on social media platforms includes the noisy

proliferation of translation systems. Machine translation tools are used to translate text, but their lack of precision can make it difficult to analyze sentiment in the translated text. Therefore, this may lead to inaccuracies and misinterpretations [14]. Fifth, analyzing the emotional inclination of online public opinion can be daunting due to the intricate nature of human emotions and the subtleties of language. Additionally, ensuring effective regulation of people's emotions on social media platforms poses a significant challenge for the Chinese government, which is important to maintain social stability [1]. Finally, one of the ways to retrieve data is to use the web language collection method, given the availability of a huge amount of available data exceeds the ability of individuals to deal with it efficiently. The process of identifying relevant data cannot be based solely on the researcher's expertise. Networking organizations can benefit from group methodology and enable a deeper understanding that can greatly enhance the research process [25].

4.4 Processing Online Content

Analyzing the sentiment of Chinese social media platforms poses several unique challenges that can make processing online content difficult. In this sub section, we will discuss two main challenges of Chinese sentiment analysis: difficulty processing in real-time and lack of information. The challenge of analyzing Chinese sentiment on social media platforms is the difficulty in processing online content and accurately distinguishing between positive and negative comments. One of the main problems is the lack of a real-time dictionary that is more contemporary and comprehensive [7]. Another big challenge is the spread of noise across social media platforms' content as advertisers sometimes use empathy harmfully. This makes it difficult to accurately fit the data distribution. In addition, the lack of information such as (age, location, gender, etc.) or the use of false information provided by many users of social media platforms further complicates the analysis process [9].

4.5 Research

Based on our reading of research and studies on Chinese sentiment analysis, we believe that researchers have had difficulty performing Chinese sentiment analysis. This is due to the difficulty of conveying and analyzing (1) emotions in Chinese, (2) the lack of content of Chinese, (3) and the lack of publicly available datasets.

4.5.1 Emotions

Chinese sentiment analysis faces several challenges that hinder progress in this field. In terms of research on emotions first, the difficulty of conveying and analyzing feelings in Chinese is a great challenge [25]. There are few studies on the mechanism of group-induced reappraisal on internet users' negative emotion regulation, which influences the accuracy of sentiment analysis models [1].

4.5.2 Chinese Content

The content is amongst the most important points in our study, it turns out that it faces challenges about the lack of content in the Chinese language, which is considered a major obstacle to research in the analysis of Chinese sentiments. Chinese sentiment resources are scarce, which limits the data available for training sentiment analysis models [2]. Despite the explosive growth of Chinese e-business and e-markets[19], there have been very few studies on sentiment analysis of Chinese content [22]. This further impedes progress in this field and the development of research and Chinese sentiment analysis methods [15].

4.5.3 Public Dataset

The SA process in the Chinese language faces challenges for researchers in terms of the nature of its data sets and the difficulty of extracting hidden feelings such as (hatred, malice, sarcasm, jokes, etc.) and this is what some researchers described as a lack of public data due to the difficulty of reaching the meaning clearly [10].

5.1 Technical

During the sentiment analysis of experimental research, researchers have faced technical challenges that can be related to either (1) the data or (2) its processing.

5.1.1 Data

Chinese sentiment analysis faces major technical challenges related to data. The accuracy of the analysis is greatly impacted by the quality of data, which is heavily dependent on the dictionary used [7]. However, dictionaries may have limitations, such as not having a large emotional lexicon with detailed classifications. Another challenge is the limited size of the data, which led to lower accuracy rates for the categories of frequent use of humor with different emoji [12, 20] and the lack of a comprehensive emotional lexicon increased the difficulty of extracting data from texts, and the use of a manual analysis method takes a lot of time and effort [8]. Earlier research has addressed the challenges of gathering data through the Internet [2] or by conducting customer surveys [19] to gain the necessary prior knowledge that businesses require to better understand their customers and their perspectives towards them [21].

5.1.2 Processing

Chinese sentiment analysis faces various technical challenges, especially in processing. One of the biggest challenges is the lack of effective processing techniques to handle the complexity of the Chinese language. With the progress of the use of machine learning methods and approaches, until this time, there have been problems in building a model that uses patterns in language to predict emotions [16]. Chinese characters are more complex than words in other languages and have different meanings depending on the context in which they are used. They are often difficult to assimilate in translation. Example (there is no Chinese equivalent for "yes") [23]. The method of translating Chinese texts into a monolingual form was used, and it was a challenge to try not to lose the original sentence contexts and semantic meanings to analyze them more accurately [14]. The presence of pictograms to express emotion may be fuzzy and may vary depending on the context, making it difficult to process and analyze [12]. In addition, detecting irony in Chinese text is difficult due to the different cultures and usage styles [4]. Sentiment analysis at the level of sentences and text faces problems in processing the text because of the analysis of multiple aspects. For example, suppose a product has both positive and negative features. In that case, it can be difficult to determine the overall sentiment of a review that discusses both aspects [19] and the use of pictorial nature to express the difficulty of processing systems [4]. On the other hand, researchers faced a challenge in processing the data, as well as the ability to measure morale and feelings directly [11]. Therefore, in order to accurately analyze sentiment in an audit, it is important to identify and analyze sentiment regarding specific aspects under review [11]. The difficulty of accessing information and its availability at the required time without delay in issuing it led to a decrease in the power and quality of the analysis and, consequently, to the loss of the public with the progress of the required data and the loss of their confidence and interest in the results [1].

6. Motivations

In our review of the literature, one of the most important sections of the study discusses and explains the importance of studying this research topic linked to Chinese SA. Based on our evaluation of previous studies, we have categorized motivations into three interrelated categories, including (1) Research, (2) Practical, and (3) Social. Details of each of these aspects are discussed below.

6.1 Research

Regarding the first part of the motivation in this paper related to research, it was meant to explain the importance of feelings in research and data sources.

6.1.1 Feelings in Research

Emotions and feelings are seen amongst the most important public manifestations of what is important to people. This shows its importance which cannot be overlooked. In SA research, feelings has been presented in different classifications, but majorly are associated with positive, negative, anger, fear, and many others [9]. Also in SA research, feelings and emotions were analyzed in different ways, especially at the word level and some at the sentence level [4]. It is widely circulated through all means of communication by various groups [5, 9]. This increases the importance of recognizing and classifying feelings significantly [26]. Amongst these examples can be seen in building automated and intelligent analytical approaches that help understand and classify feelings and emotions [12], not only in normal cases but also in hidden emotions [19, 27], which in turn can assist and aid in better understanding the Chinese language given the influence of its semantic units and characteristics [16, 22].

6.1.2 Data Sources

As previously mentioned with regard to the importance of feelings-related data in various research venues, there are many platforms for getting feelings and emotions data, including social media platforms [25], blogs or questionnaires [28], or by analyzing previous literature [10]. Therefore, this shows the importance of establishing and providing a database for Chinese feelings [16]. This information extraction can help facilitate many analytical benefits for different sectors, like industry and academics. [13]. Another significant factor in data sources is establishing dedicated dictionaries for Chinese language SA research. Based on an analysis of previous academic works, creating or developing a Chinese dictionary can help predict people's reaction, which will help improve many aspects of life [23]. Amongst the good practices to motivate research in this area is by integrating some of the current Emoji dictionaries with common words [6] and using them for many industrial applications, for example, predicting fluctuations in economic matters to help investors understand the market better [6].

6.2 Practical

Based on previous literature analysis, it turns out that there are a lot of applications, experiments, and support in SA to present proper methods that facilitate the translation and understanding of feelings within the Chinese language. Therefore, this sub-section discusses practical motivations with respect to (1) Support the process, (2) experiments, methods, and analytical algorithms.

6.2.1 Support the Process

According to previous studies, researchers have stressed the importance of developing and supporting methods to improve performance and developing approaches, techniques, and algorithms that enable the classification and analysis of feelings in the Chinese language more

efficiently and effectively [7, 10, 19, 28]. Current and future studies need to reinforce learning principles to study various aspects of the Chinese language due to the different phonetic, textual, orthographic features and other characteristics [2]. Studies also need to integrate these approaches to enhance the performance of analysis [15], in addition to other types of support that help address extraction problems due to insufficient interaction and difficulty in categorizing and analyzing sentiments [14, 19], dealing with Chinese language grammar [22, 24], and solving the challenges of texts containing symbols and long phrases [14, 24]. Some researchers also pointed out the use of the most widely used social platforms to benefit from the availability of their data. For instance, in the study [12], the authors used Weibo, a Chinese website for microblogging and one of China's largest social networking platforms. Amongst the other practical motivations is researchers' utilization of the fields of artificial intelligence and deep learning [7].

6.2.2 Experiments, methods, and Analytical Algorithms

SA can be applied in different languages. Each language has its own characteristics, so the type and method of analysis and algorithms differ. As indicated in previous studies, using algorithms to find "Beginning Definite Base Noun Phrases" in the Chinese language was characterized by accuracy and was considered difficult to retrieve because of the characteristics of the Chinese language that differ from other languages [13]. It was also mentioned that the attempt to explore public opinion regarding items like products is one of the concerns that got the attention of the researchers [8, 13]. Recently, a unified classification framework based on CNN long texts [24], microblogs [16], and word2vec [20] has been proposed to improve the accuracy and classification performance of sentiments in Chinese [12]. Previous studies also encouraged utilizing approaches to reach effective outcomes [4] and the importance of knowing and defining feelings from a more emotional aspect [26]. Some researchers confirmed that the use of a radical-based approach has an effective role in excavating texts in Chinese [21], especially the general opinion, which is classified into three levels: first document, second sentence, and third feature [13], which in turn will help in reaching accurate analysis of feelings [27]. Some studies sought to design an algorithm that presents an overall high quality [5] and combine it with approaches and techniques that help remove ambiguity from Chinese language texts and create new methods that facilitate understanding of these feelings [15]. Another study used a hybrid approach to improve the performance of classification and information transmission through NLP [14]. This shows the importance of using artificial intelligence, which is currently popular among researchers, as it helps save time and effort and increases its ability to perform complex tasks that help analyze feelings in the Chinese language [11].

6.3 Social Media Platforms

It is worth noting that at present, the development of technology and the availability of social platforms are important factors in the development of the world and the prosperity or failure of many sectors and sometimes countries. It is considered the fastest way to express opinions and feelings through the various platforms available to different cultures, so it will be discussed concerning social media platforms and their impact on (1) social life, (2) economy, and (3) Decision-making.

6.3.1 Social Media with Social Life

The effect of SA through social media is the information retrieval process if people's feelings are expressed about a particular issue [9]. It can be said that feelings are the key to everything that may confront us [2]. Through public opinion, people's feelings can be detected [5] in different contexts, and an analytical attempt is performed to analyze and reduce potential risks anticipated to predict

people's feelings after studying them [27], especially in cases of disasters and crises. After that, an attempt is required to control them by spreading awareness and educational awareness that helps them overcome unwanted feelings [1].

6.3.2 Social Media with Economy

Social media SA is the process of gathering and analyzing information about how people talk about your brand on social media. Instead of counting comments and mentions, sentiment analysis considers emotions and opinions expressed in these comments. This can help business owners get to know their customers more [19] and predict their economic growth among their competitors alongside their strength in the markets [22]. In addition, companies are certainly interested in their image with their investors and those who are expected to invest with them, and studying feelings in the Chinese language can help them to know and improve their position [6, 11]; this can be applicable in various ways by increasing advertisements and promoting or modifying a product that will be more accepted by the public [28].

6.3.3 Social Media Making Decision

People use social media to share their experiences, reviews, information, tips, warnings, and/or any issues concerning their "connection" or their friends. This information is a useful resource that may influence decision-making. It is a double-edged sword, as it impacts the convergence of these feelings of different polarities and the cause of these feelings. From an economic point of view, SA helps the factory or supplier to predict consumer behaviour in advance [25] and learn about the performance of its products or services, which helps decision-makers to develop and improve its products or services [8] and increases investment opportunities in this company by improving their image in front of investors [7]. This makes it easier for them to understand and make a decision to conclude contracts with companies [6]. Also, social media platforms play an important role in consumers' knowing the product through previous experiences and the opinions of other users [28]. Different companies were keen to establish their own websites, making it easier for their customers to express their opinions; this plays an important role in developing and improving it and making more reliable service, as is the case with the well-known website (Alibaba), which is a provider of e-commerce and technological infrastructure services [21].

7. Recommendations

The researchers indicated the importance of integrating and developing SA solutions for studies in the Chinese language, as it is the second language that is exchanged globally after English [10]. Researchers recommended the importance of conducting future studies on the importance of analyzing neutral feelings, in addition to positive and negative ones [1, 3, 22]; this includes the importance of expanding the Data source by integrating images, emojis, videos, audio clips, media and public opinion with social media posts [3, 15], creating an appropriate processing method to help classify and sort out the repetitive words and unhelpful links which help to reach feelings more clearly. This makes it easier for companies to make decisions in several sectors [5, 12]. This also helps improve the performance of their products and services [8]. In addition, some research works suggested the use of semantic meaning and cognitive representation with data containing small skills and unbalanced data [2, 9]. It has also been suggested that it be used with the writing system in order to improve understanding of meaning in sentiment analysis [19]. The researchers mentioned focusing on using supervised methods, especially with big data [10]; also, using the (ontology and radical) based approach would help the comparison more efficiently [21]. Therefore, this enables us to create a more comprehensive emotional lexicon to classify feelings and adopt it for subsequent studies [28].

Others have shed light on technical recommendations. In that regard, technical-related recommendations involve those that discuss implementing new modules. Some researchers discuss the importance of microblogging mining [16]. Some studies have suggested integrating knowledge with deep learning, as it helps to understand and classify data [24], which facilitates the ability to identify investor surfers' feelings and predict [4, 7]. Others have recommended using new indicators that help in sentiment analysis, such as (investor-sentiment-indicator-ISI) in prospective studies [11]. In addition to that, researchers reported the importance of integrating knowledge and dictionary-based methods with machine learning to analyze sentiment [6]; this can help in utilizing the big data for further studies [22]. Scientists emphasized the possibility of including robots in sentiment analysis by analyzing face impressions [20]. Other researchers indicated the importance of increasing studies due to the large number of future work fields for studying Chinese sentiment analysis, such as medicine, transportation, and others [10], and also developing and integrating algorithms that were used in previous studies [22, 27]. They also encouraged building new groups, such as merging different languages with the Chinese language [19], futuristic [14], and taking into account contextual variables [26].

8. Conclusion

This paper presents a systematic review and analysis of the literature on sentiment analysis in Chinese, gaining a comprehensive understanding of sentiment analysis based on user-generated data and providing a comprehensive overview of the challenges, motivations, and recommendations for conducting sentiment analysis in Chinese. Through the use of a systematic review protocol, we were able to identify key findings from a variety of sources, including academic journals, conference proceedings, and online forums. Our analysis revealed that although significant progress has been made in the development of Chinese sentiment analysis tools, there are still challenges to be overcome, such as the lack of standardized data sets and the difficulty in dealing with the nuances of the Chinese language. Based on our review, we recommend that future research focus on addressing these challenges and developing more accurate and reliable sentiment analysis tools for the Chinese.

Author Contributions

Conceptualization, methodology, software, validation, formal analysis, investigation, S.A.M., and I.A.; resources, S.A.M.; data curation, I.A.; writing—original draft preparation, writing—review and editing, supervision, S.A.M., and I.A. All authors have read and agreed to the published version of the manuscript.

Funding

This research received no external funding.

Data Availability Statement

Data will be made available on request.

Conflicts of Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

This research was not funded by any grant.

References

- [1] Zhang, W., et al. (2022). CNN-LSTM neural network model for fine-grained negative emotion computing in emergencies. *Alexandria Engineering Journal*, 61(9), 6755-6767. <https://doi.org/10.1016/j.aej.2021.12.018>
- [2] Peng, H., Cambria, E., & Hussain, A. (2017). A review of sentiment analysis research in Chinese language. *Cognitive Computation*, 9(4), 423-435. <https://doi.org/10.1007/s12559-017-9470-8>
- [3] Hao, J.-X., et al. (2020). Introducing news media sentiment analytics to residents' attitudes research. *Journal of Travel Research*, 59(8), 1353-1369. <https://doi.org/10.1177/0047287519884657>
- [4] Wen, Z., et al. (2022). Sememe knowledge and auxiliary information enhanced approach for sarcasm detection. *Information Processing & Management*, 59(3), 102883. <https://doi.org/10.1016/j.ipm.2022.102883>
- [5] Wang, G., et al. (2019). Studies on a multidimensional public opinion network model and its topic detection algorithm. *Information Processing & Management*, 56(3), 584-608. <https://doi.org/10.1016/j.ipm.2019.02.005>
- [6] Zhang, W., et al. (2021). Predicting stock market volatility based on textual sentiment: A nonlinear analysis. *Journal of Forecasting*, 40(8), 1479-1500. <https://doi.org/10.1002/for.2777>
- [7] Jiang, Z., et al. (2022). Investor sentiment and machine learning: Predicting the price of China's crude oil futures market. *Energy*, 247, 123471. <https://doi.org/10.1016/j.energy.2022.123471>
- [8] Wei, G., Cai, W., & Qin, L. (2016). Research on product feature extraction for Chinese reviews. In 2016 International Conference on Logistics, Informatics and Service Sciences (LISS). IEEE. <https://doi.org/10.1109/LISS.2016.7854478>
- [9] Zhang, Y., et al. (2020). Sentiment analysis of sina Weibo users under the impact of super typhoon lekima using natural language processing tools: A multi-tags case study. *Procedia Computer Science*, 174, 478-490. <https://doi.org/10.1016/j.procs.2020.06.114>
- [10] Tubishat, M., Idris, N., & Abushariah, M.A. (2018). Implicit aspect extraction in sentiment analysis: Review, taxonomy, opportunities, and open challenges. *Information Processing & Management*, 54(4), 545-563. <https://doi.org/10.1016/j.ipm.2018.03.002>
- [11] Ruan, Q., et al. (2020). A new investor sentiment indicator (ISI) based on artificial intelligence: A powerful return predictor in China. *Economic Modelling*, 88, 47-58. <https://doi.org/10.1016/j.econmod.2019.09.010>
- [12] Li, D., et al. (2020). HEMOS: A novel deep learning-based fine-grained humor detecting method for sentiment analysis of social media. *Information Processing & Management*, 57(6), 102290. <https://doi.org/10.1016/j.ipm.2020.102290>
- [13] Zhu, S., et al. (2009). Research on feature extraction from Chinese text for opinion mining. In 2009 International Conference on Asian Language Processing. IEEE. <https://doi.org/10.1109/IALP.2009.53>
- [14] Zhu, X., et al. (2022). Leveraging bilingual-view parallel translation for code-switched emotion detection with adversarial dual-channel encoder. *Knowledge-Based Systems*, 235, 107436. <https://doi.org/10.1016/j.knosys.2021.107436>
- [15] Peng, H., et al. (2021). Phonetic-enriched text representation for Chinese sentiment analysis with reinforcement learning. *Information Fusion*, 70, 88-99. <https://doi.org/10.1016/j.inffus.2020.12.004>
- [16] Xu, D., et al. (2020). Deep learning based emotion analysis of microblog texts. *Information Fusion*, 64, 1-11. <https://doi.org/10.1016/j.inffus.2020.06.007>
- [17] Alamoodi, A.H., et al. (2021). Sentiment analysis and its applications in fighting COVID-19 and infectious diseases: A systematic review. *Expert Systems with Applications*, 167, 114155. <https://doi.org/10.1016/j.eswa.2020.114155>
- [18] Xu, Q.A., Chang, V., & Jayne, C. (2022). A systematic review of social media-based sentiment analysis: Emerging trends and challenges. *Decision Analytics Journal*, 3, 100073. <https://doi.org/10.1016/j.dajour.2022.100073>
- [19] Xiao, Y., et al. (2022). Towards Lean Automation: Fine-Grained sentiment analysis for customer value identification. *Computers & Industrial Engineering*, 169, 108186. <https://doi.org/10.1016/j.cie.2022.108186>
- [20] Jia, K. (2021). Chinese sentiment classification based on Word2vec and vector arithmetic in human-robot conversation. *Computers and Electrical Engineering*, 95, 107423. <https://doi.org/10.1016/j.compeleceng.2021.107423>
- [21] Chao, A.F. & Yang, H.-L. (2018). Using Chinese radical parts for sentiment analysis and domain-dependent seed set extraction. *Computer Speech & Language*, 47, 194-213. <https://doi.org/10.1016/j.csl.2017.07.002>
- [22] Zhang, C., et al. (2009). Sentiment analysis of Chinese documents: From sentence to document level. *Journal of the American Society for Information Science and Technology*, 60(12), 2474-2487. <https://doi.org/10.1002/asi.21206>
- [23] Du, Z., et al. (2022). Language and Domain Specificity: A Chinese Financial Sentiment Dictionary. *Review of Finance*, 26(3), 673-719. <https://doi.org/10.1093/rof/rfab034>
- [24] Zhang, C., et al. (2021). Grammar guided embedding based Chinese long text sentiment classification. *Concurrency and Computation: Practice and Experience*, 33(21), e6439. <https://doi.org/10.1002/cpe.6439>

- [25] Yuntao, Z. (2021). Research on Feature Extraction and Tendency Analysis of Chinese Language Sentimental Expression in Network Context. In 2021 International Conference on Intelligent Transportation, Big Data & Smart City (ICITBS). IEEE. <https://doi.org/10.1109/ICITBS53129.2021.00096>
- [26] Li, J. & Ren, F. (2013). A hybrid approach for word emotion recognition. *IEEJ Transactions on Electrical and Electronic Engineering*, 8(6), 616-626. <https://doi.org/10.1002/tee.21905>
- [27] Dai, Z., Dai, W., Liu, Z., Rao, F., Chen, H., Zhang, G., ... & Liu, J. (2019, September). Multi-task multi-head attention memory network for fine-grained sentiment analysis. In *CCF International Conference on Natural Language Processing and Chinese Computing* (pp. 609-620). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-32233-5_47