Journal of King Saud University - Computer and Information Sciences 36 (2024) 102048

Contents lists available at ScienceDirect

Journal of King Saud University - Computer and Information Sciences

journal homepage: www.sciencedirect.com

Review Article

Sentiment analysis methods, applications, and challenges: A systematic literature review

Yanying Mao [a,b], Qun Liu [a,*], Yu Zhang [c]

[a] Chongqing Key Laboratory of Computational Intelligence, Key Laboratory of Big Data Intelligent Computing, College of Computer Science and Technology, Chongqing University of Posts and Telecommunications, Chongqing 400065, China
[b] Department of Communication Engineering, Chongqing College of Electronic Engineering, Chongqing 401331, China
[c] School of Economics and Management, Chongqing Jiaotong University, Chongqing 400074, China

ARTICLE INFO

Keywords:
Sentiment analysis
Methods
Applications
Large language models
Challenges

ABSTRACT

With the expansion of Internet-based applications, the number of comments shows explosive growth. Analyzing the attitudes and emotions behind comments provides powerful assistance for businesses, governments, and scholars. However, it is hard to effectively extract user's attitude from the massive amounts of comments. Sentiment analysis (SA) provides an automatic, fast and efficient tool to identify reviewers' opinions and sentiments. However, the existing literature reviews cover a limited number of studies or have a narrow field of studies for sentiment analysis. This paper provided a systematic literature review of sentiment analysis methods, applications, and challenges. This systematic literature review gives insights into the goal of the sentiment analysis task, offers comparisons of different techniques, investigates the application domains of sentiment analysis, highlights the challenges and limitations encountered by scholars, provides suggestions on possible solutions and explores the future research directions. The study's findings emphasize the significant role of artificial intelligence technologies in automatic text sentiment analysis and highlight the importance of sentiment analysis in people's production and life. This research not only contributes to the existing sentiment analysis knowledge body but also provides references to scholars and practitioners in choosing a suitable methodology and good practices to perform sentiment analysis.

1. Introduction

In the last decade, the rapid development of the Internet has generated a huge amount of online comments, and analyzing the sentiments of netizens in comments is of great significance for social stability and development (Birjali et al., 2021). In response to this trend, SA technology has emerged. SA, also known as opinion analysis or opinion mining, is an important study area in NLP, which is designed to extract and analyze sentiment and views from text automatically (Chaturvedi et al., 2018; Liu and Zhang, 2012). SA is crucial for the development of artificial intelligence (Liu and Zhang, 2012).

The types of sentiment we can find include positive, neutral and negative, and can be further divided into surprise, trust, anticipation, anger, fear, sadness, disgust, joy and so on (Bose et al., 2020). From a language perspective, sentiment analysis research can use various types of natural languages, such as Chinese (Peng et al., 2017), English (Rodríguez-Ibánez et al., 2023), Arabic (Oueslati et al., 2020), German (Remus et al., 2010), Portuguese (Pereira, 2021), Russian (Smetanin, 2020), Spanish (Osorio Angel et al., 2021), etc.

Machines can only make intelligent responses by analyzing and understanding human emotional expressions, thus better serving humanity. For example, sentiment analysis is of great importance in supporting the Human Machine Intelligence Q&A (Eskandari et al., 2015) and the epoch-making large language models (LLM), i.e. ChatGPT and ERNIR (Huang et al., 2022b; Sudirjo et al., 2023; Susnjak, 2024). In addition, SA plays a necessary role in the development of Internet social platforms, e-commerce, finance, healthcare, and government sectors (Rodríguez-Ibánez et al., 2023).

As the Internet occupies most of people's lives, online comments are exploding (Jain et al., 2021). Analyzing tens of thousands of comments and capturing the psychological thoughts of commentators efficiently, quickly, and low-cost automatically, can aid public opinion supervision and decision-making in various industries.

Therefore, SA has become increasingly popular among research

* Corresponding author.
E-mail addresses: maoyanying@foxmail.com (Y. Mao), liuqun@cqupt.edu.cn (Q. Liu), zhangyu_0112@foxmail.com (Y. Zhang).

https://doi.org/10.1016/j.jksuci.2024.102048
Received 24 January 2024; Received in revised form 12 April 2024; Accepted 29 April 2024
Available online 30 April 2024

communities recently, with the increase in the number of related research work and publications. In the past decade, SA has attracted the attention of more and more scholars, as the number of papers focusing on SA has increased significantly. Fig. 1fig1 depicts the rising time distribution of articles in Web of Science.

However, it is noteworthy that such a volume of sentiment analysis publications makes researchers difficult to quickly and efficiently obtain critical and relevant research. To address this problem, researchers argued that a systematic and comprehensive literature review is an effective way. For instance, Birjali et al. (2021) presented a review of SA approaches, challenges and trends. This paper described the preprocessing process and categorized various SA approaches. Some reviews of SA focused on deep learning (DL) methods, such as (Mercha and Benbrahim, 2023; Yadav and Vishwakarma, 2020). The survey of SA on lexicon-based approaches compared the SA tools, including Natural Language Toolkit (NLTK), Text blob, and Valence Aware Dictionary and Esntiment Reasone (VADER), to find that VADER outperforms other tools (Bonta and Janardhan, 2019). Some reviews of SA have been conducted in specific fields. The survey by Shaik et al. (2023) reviewed the sentiment annotation techniques, sentiment analysis tools and applications in education data. Discussing the impact and challenges of SA in education. A review explicitly focused on SA applications related to health was provided by Zunic et al. (2020). Xu et al. (2022b) explored the challenge and compared different SA studies related to the methods, text languages, datasets and evaluate metrics in social media. Other SA reviews focus on specific levels of text, such as aspect-level sentiment analysis (Do et al., 2019; Phan et al., 2023; Zhang et al., 2022a) and document-level (Alshuwaier et al., 2022; Behdenna et al., 2018). Chan et al. (2023) discussed the development of sequential transfer learning and pre-trained models, reviewed the application and suggested future research directions. In addition to English, there are also many reviews of SA in Arabic (Alyami et al., 2022; Matrane et al., 2023). The prompt-based methods for SA have been summarized and the large-scale language models has been discussed in (Bu et al., 2023). In addition, multimodal SA about technologies and challenges are the future research direction (Das and Singh, 2023; Gandhi et al., 2023).

However, as far as we know, the existing surveys focus on special sub-fields of SA, such as ABSA and skip some of the latest SA technologies due to these surveys concentrate only on lexicon-based methods, machine learning and transformer learning. Although this work also discussed these tasks, it varies from the prior work in that it covers the most approaches. In addition, the existing surveys study SA from a particular domain application, such as e-commerce. For example, the review presented by Huang et al. (2023) aimed to examine techniques and explore the future directions of SA in e-commerce. And the review summarized the techniques and algorithms of SA in e-commerce (Marong et al., 2020). This survey differs from the previous studies by providing an up-to-date perspective on most applications and large language models. Furthermore, previous reviews lacked sufficient literature and incompletely described challenges and future trends. To fill this gap, this study covers a wider range of literature, which studies more than 200 research papers. And the survey utilized manual analysis for deeper insights.

The paper makes valuable contributions by presenting a comprehensive survey of SA as it discusses this topic from all aspects, including a wider range of technologies, cutting-edge applications and comprehensive challenges. Its strengths lie in the wide coverage of literature, manual analysis for deeper insights, and an up-to-date perspective on applications and large language models. This article is very beneficial for beginners and researchers as it enables them to gain rich knowledge about the field in one study. The significant contributions of this survey can be summarized as follows:

- A wide range of researches were manually studied to give deeper insights for scholars to construct methodologies based on prior researches.
- Comparisons of current sentiment analysis methods in order to assist researchers in selecting appropriate approaches to perform this task.
- An up-to-date perspective on applications and large language models to keep up with the latest research developments.
- A comprehensive summary of SA challenges and potential solutions for references in order to identify where best to focus future research efforts.

The remainder of this paper is organized as follows. Section 2 introduces the review methodology. Section 3 classifies the sentiment analysis levels. Section 4 discusses and categorized the sentiment analysis methods. Section 5 shows the datasets and evaluation for SA. Section 6 describes the application domains of SA. Section 7 discusses and proposes the challenges of SA.

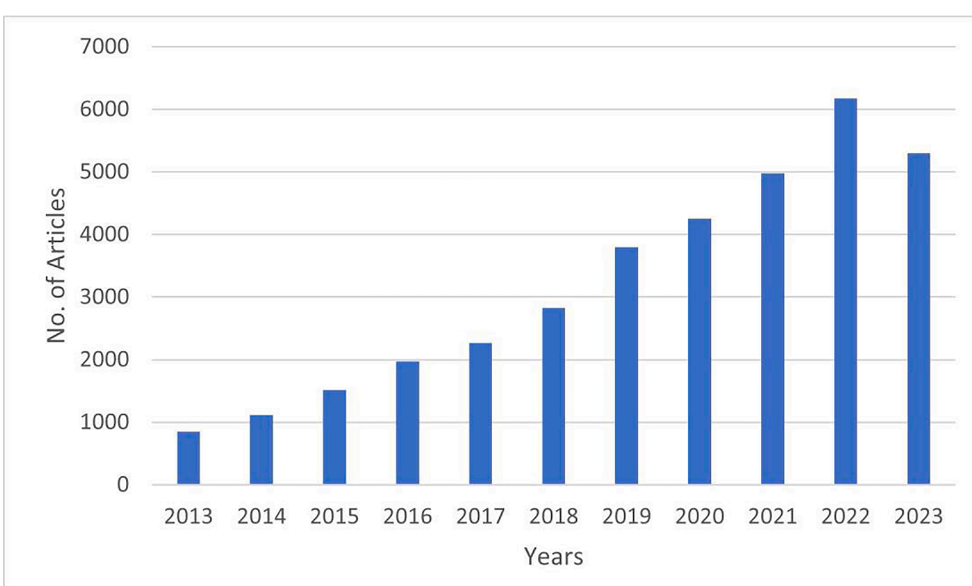

Fig. 1. Annual distribution of the articles amount in Web of Science.

2. Systematic review methodology

Systematic literature review (SLR) is a literature review strategy that follows a systematic process to identify, collect, select and analyze primary studies on a particular topic. SLR provides an unbiased, detailed, and fast overview of current studies. This review is based on the Preferred Reporting Items for the Systematic Review and Meta-Analysis (PRISMA) framework to carry out (Sarkis-Onofre et al., 2021). The review process comprises four steps: definition of research questions, literature search, screening data and data analysis.

2.1. Research questions

There are many factors that affect the performance of SA. These factors involve the level of SA, models and techniques, datasets, evaluation metrics, application domains and so on. Considering these factors, we define the research questions as follows:

RQ1: What are the methods used to conduct SA?
RQ2: What are the most effective methods of SA?
RQ3: What datasets and evaluation metrics are used in SA?
RQ4: What are the most popular application areas of SA?
RQ5: What are the challenges and future research direction of SA?

2.2. Search strategy

The literature search followed the PRISMA guidelines, which contain three stages: identification stage, screening stage and inclusion stage. The flow for literature selection is depicted in Fig. 2.

In the identification stage, the review searches for journal papers and conference proceedings from scientific digital databases and academic search engines from 2013 until 2023. These databases include Scopus, Web of Science, Elsevier Science Direct, IEEE/IET Electronic Library, ACM digital library, Springer Link, and Google Scholar. The keywords used in the search process are as follows: sentiment analysis, method, application and challenge. Research expressions employ logical operator AND to combine the keywords to conduct the inquiries. The search indices are limited to titles, abstracts, and keywords.

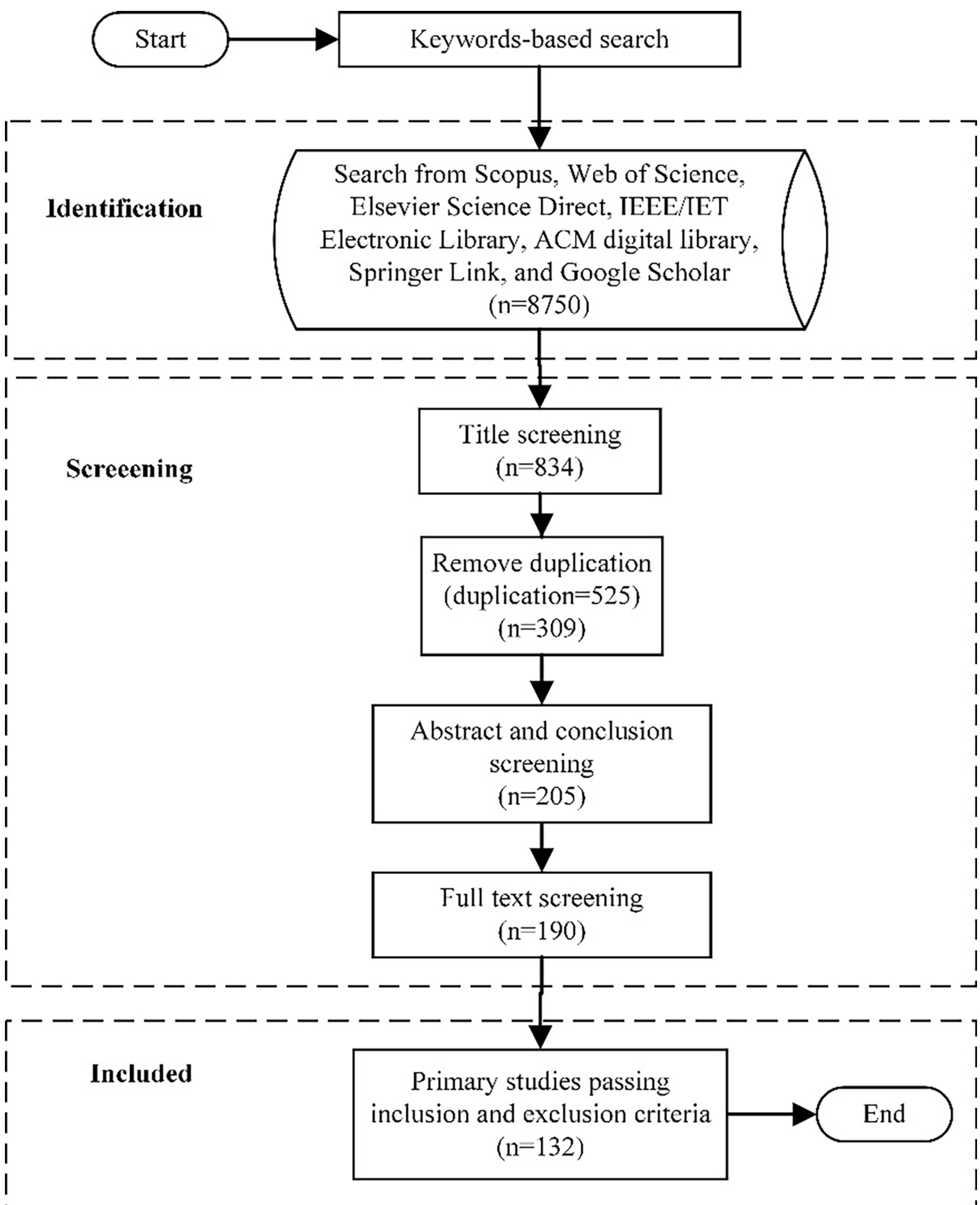

Fig. 2. The PRISMA flow diagram for literature selection.

2.3. Inclusion and exclusion criteria

In the screening stage, inclusion and exclusion criteria are employed to determine whether the primary studies meet certain criteria. If the paper satisfies all the exclusion criteria, it is included. However, if it meets at least one exclusion criteria, it is excluded. The inclusion and exclusion criteria we applied are shown in Table 1.

Taking the criteria into account, we scan the title and abstract of the study to screen suitable articles. If the title and abstract are inconclusive, we screen by studying the full text.

As a result, 132 articles were finally included for further detailed analysis in the inclusion stage.

2.4. Data analysis

In this step, we discuss the summary information from primary studies in the inclusion phase. We extract two types of data from each study. One data type is information about the author, publication years and publisher. The other type is the data that focuses on the research questions of our SLR. These include methods/techniques, application domain, public datasets, evaluation metrics, and existing challenges, which are significant attributions to discuss in the study of SLR.

2.5. Research contents

This survey mainly contains three parts, namely methods, applications, and challenges, as shown in Fig. 3. Firstly, we have described and made a taxonomy for SA methods. Secondly, this study has researched various publications to investigate sentiment analysis application areas. Finally, a discussion and challenges section has been created to analyze, criticize, and evaluate the reviewed studies with highlighting the future directions and research gaps. The methods, applications, discussion and challenges of SA are reviewed in the subsequent sections.

3. Sentiment analysis levels

The scope of SA can be mostly divided into the document, phrase, and aspect levels according to the text range (Behdenna et al., 2016; Do et al., 2019). Section 3.4 defines the types of sentiment we can find and differentiates between the analyzed expressions in terms of language and type.

3.1. Document-level SA

This task is researching the emotion of the entire document. Document-level SA treats each document as an independent object, and a document has only one sentiment polarity, so this task is coarse-grained (Behdenna et al., 2018). Mao et al. (2022) conducted document-level SA using attention-based bi-directional LSTM network and CNN. Wen et al. (2020) proposed a speculative SA model and they speculated that similar reviews are more likely to be written by users with similar sentiments. So, they utilized similar documents to improve SA accuracy.

Table 1
The inclusion and exclusion criteria.

Inclusion Criteria	Exclusion Criteria
The paper published between 2012 and 2023	The study not focuses on SA
The paper written in English	The paper not written in English
The research focuses on research questions of our SLR	The research not focuses on any research questions
The paper published in conference or journal	The datasets in the form of images, audios, videos, instead of text

3.2. Sentence-level SA

Sentence-level SA aims to classify the sentiment polarity of a sentence into positive, negative, or neutral (Liu, 2012). This task categorized each sentence into objective sentences and subjective sentences. An objective sentence was referred to a sentence that does not convey any opinion. Subjective sentences describe the reviewer's thoughts and ideas, so we can divide subjective sentences into positive or negative sentiments. The paper used a sequence model to categorize sentences' sentiment polarity based on their target (Chen et al., 2017). Su et al. (2023) proposed a supervised sentence-level SA method based on gradual machine learning. This study exploited linguistic features of Urdu language to conduct this task (Altaf et al., 2023).

3.3. Aspect-level SA

Aspect-level SA is a finer analysis, and the algorithm mainly aims to model the relationship among the aspect term, aspect category, opinion term, and sentiment polarity (Wu et al., 2018). Generally speaking, aspect-level sentiment analysis requires two steps: first, extracting the aspect terms and opinion terms, then identifying the sentiment polarity of aspects. Considering the sentence, "This restaurant's steak is delicious." The "steak" is an aspect term of the aspect category "food", "delicious" is the opinion term and the sentiment polarity of "steak" is positive. Zhang et al. (2021) developed a text generation method to do aspect-level SA and the performance is better than the classification method and MLM. The study in (Huang et al., 2022a) utilized aspect-specific contextual location information to assign different weights and reduced the error in the judgment of sentimental polarity. Zhu et al. (2024) constructed a deformable CNN to incorporates a cross-correlation attention and contextual phrases for aspect-level SA.

4. Sentiment analysis techniques

Generally, SA approaches can be divided into four categories: lexicon-based, traditional machine-learning, deep-learning and hybrid approaches (Madhoushi et al., 2015; Sankar and Subramaniyaswamy, 2017; Thakkar and Patel, 2015). Sentiment analysis approaches are depicted in Fig. 4. In addition, researchers have been searching for better methods to complete the task more accurately and reduce computer expenses.

4.1. Lexicon-based approach

This method needs utilize a sentiment lexicon that assigns scores to the collected tokens (Bonta and Janardhan, 2019; Taboada et al., 2011). It is an unsupervised technique, and this method is domain reliance because the same word in different fields might have different sentiments. For example, two sentences, "The project is taking too long" and "Her fingers are long and slender". The term "long" in the first statement is negative because people like short project times; however, it is positive in the second sentence since the finger is longer, the finger is more attractive. We can overcome this problem by adopting a dictionary adaption technique. The lexicon-based approaches are usually divided into two methods: dictionary-based and corpus-based methods (Mitra, 2020). Table 2 lists the disadvantages and benefits of lexicon-based techniques.

4.2. Machine learning approach

4.2.1. Conventional machine learning classifiers

Machine learning methods separated datasets into train datasets and test datasets (Agarwal et al., 2016; Boiy and Moens, 2009; Mitra, 2020). Machine learning models can learn a lot of known information from the train datasets. The test datasets are then used to analyze the effect of models. Conventional machine learning classifiers, such as Naive Bayes

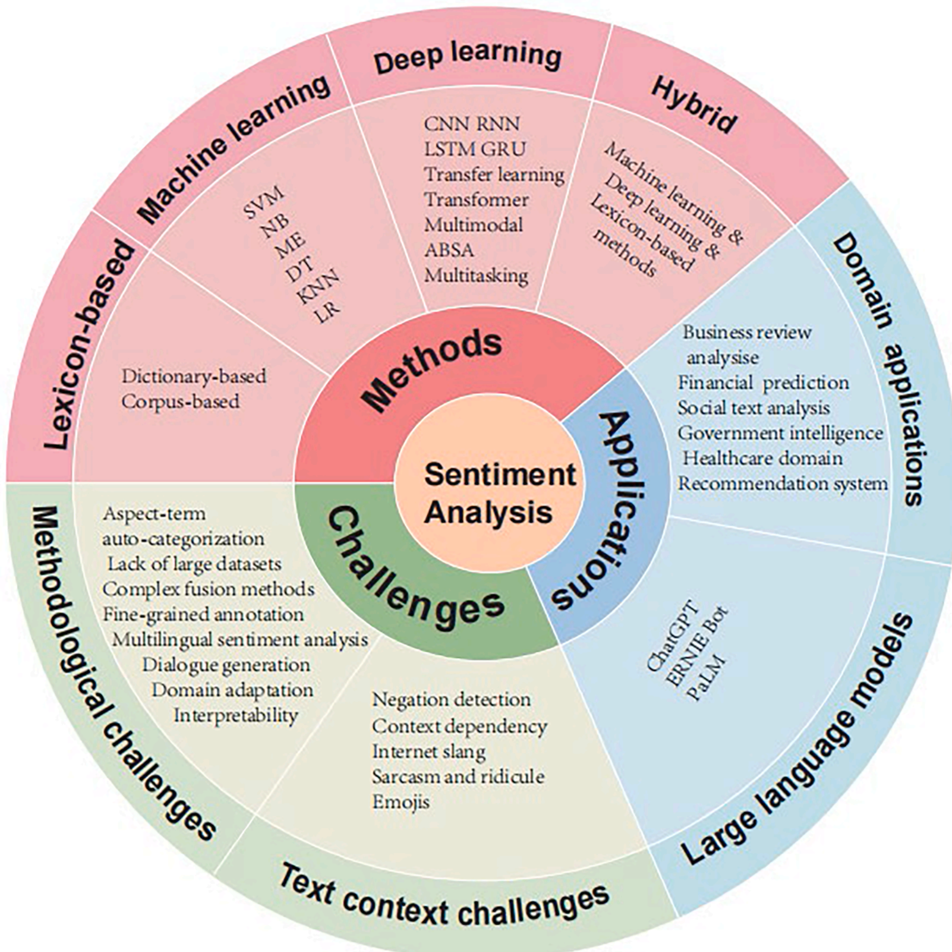

Fig. 3. Research contents.

(NB) (Kang et al., 2012), support vector machines (SVM) (Ahmad et al., 2017), decision trees (DT) (Myles et al., 2004), may be employed for sentiment classification, each having advantages and downsides.

The SVM's ability to distinguish between different hyperplanes helps maximize class boundaries (Khoshnevisan et al., 2015). It has a strong theoretical foundation, and, in many circumstances, it performs more correctly than most other machine learning algorithms (Cortes and Vapnik, 1995; Moraes et al., 2013). NB is a probabilistic classification strategy based on the Bayes theorem to describe an event's probability based on prior knowledge (Mubarok et al., 2017). The aim of maximum entropy (ME) is to make the classification result have maximum information entropy under the premise of satisfying the constraint conditions (Nigam et al., 1999; Wu, 2012). The DT algorithm uses training samples to build a tree for text SA (Kingsford and Salzberg, 2008). K-nearest neighbors (KNN) classified test data to compare with neighbors (Daeli and Adiwijaya, 2020). The K value is a hyper-parameter that can be chosen in algorithms such as Grid search and Randomized search cross-validation. Logistic regression (LR) can predict categories and obtain approximate probability predictions. However, LR is prone to underfitting and has low classification accuracy (Kleinbaum et al., 2002). Table 3 shows the benefits and drawbacks of various approach summary evaluations of sentiment analysis methods.

4.2.2. Deep learning-based classifiers

The convolutional neural network (CNN) is a feed-forward neural network with convolutional computation and pooling operation (Chen, 2015; Li et al., 2021b). CNN originated in the computer vision field and then extended to many fields, such as NLP. Chen (2015) proposed a famous CNN sentiment analysis method built on word2vec for sentence-level sentiment categorization. This method outperformed rival approaches. This paper proves the practicability of pre-training words embedded in deep learning.

The recurrent neural network (RNN) (Liu et al., 2016) was widely used in SA because it guarantee that information about a long sequence was captured and remembered (Zaremba et al., 2014). The most outstanding benefit of RNN is that it uses prior knowledge and thus can remember the previous information. Long short-term memory (LSTM) networks is a special kind of RNN, which can relief gradient explosion and vanishing problems (Hochreiter and Schmidhuber, 1997). Besides, LSTM incorporates a gating mechanism to solve long-distance dependency that ordinary RNN cannot solve. A cached LSTM (Xu et al., 2016) can store information far from the current place in a sequence. Tai et al. (2015) developed the Tree-LSTM that was superior to various LSTM baseline approaches. The structures of gated recurrent unit (GRU) networks and LSTM are similar, and the effects are almost good. The advantage of GRU is that the model is simpler and the convergence speed is faster. But when the amount of data is large, LSTM works better because there are more gates and more parameters (Abdelgwad et al., 2022).

4.3. Hybrid approach

The hybrid approach combines lexicon-based, traditional machine learning and deep-learning approaches (Appel et al., 2016). In this

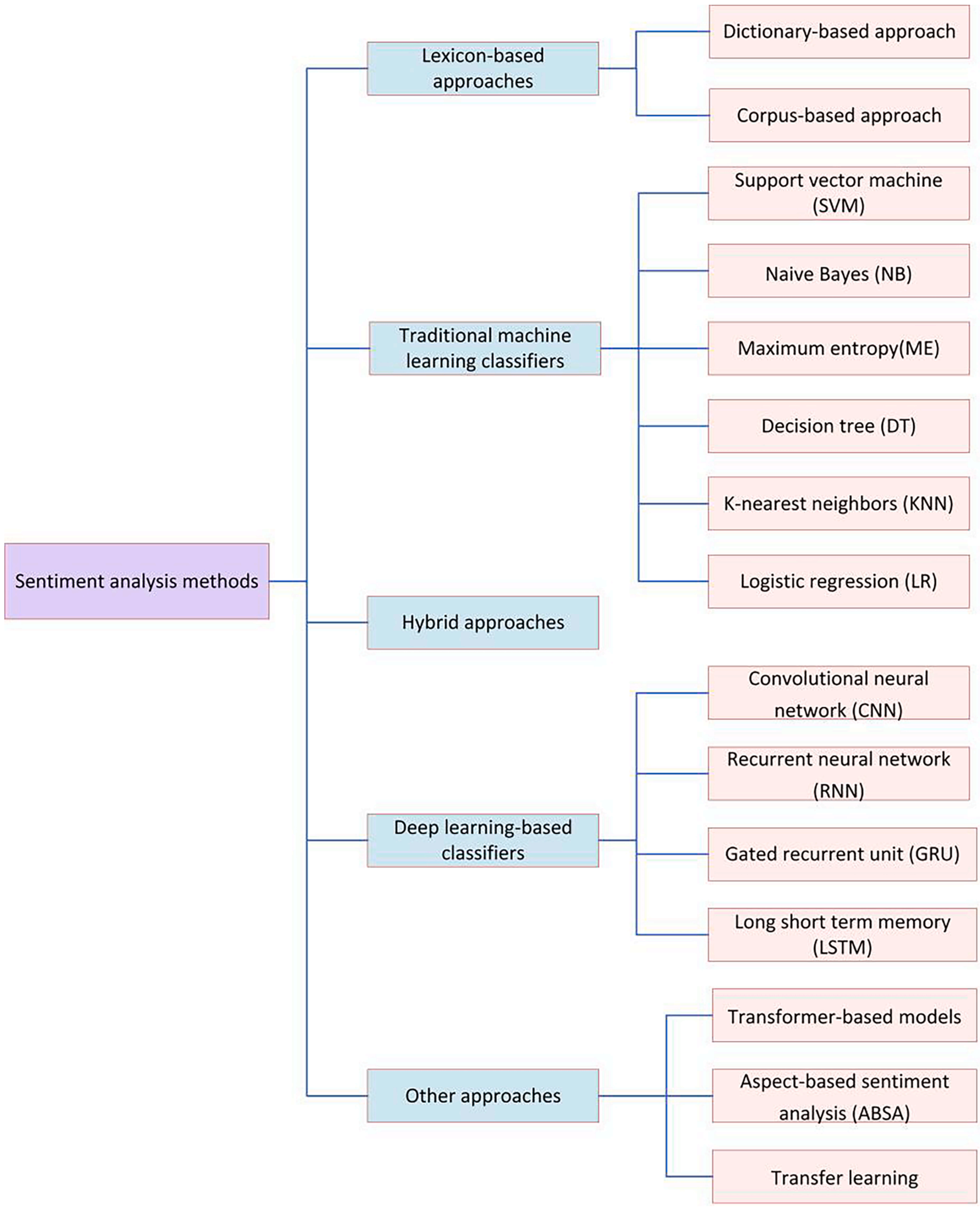

Fig. 4. Sentiment analysis Approaches.

popular fashion, sentiment analysis combines linguistic analysis and contextual semantics of words. Chang et al. (2020) exploited a hybrid technique combined SVM and Relief algorithms. To train data, they use as many as 6900 tweets to train the model. Their model outperforms most models when 96 percent of the features are included. The study also concluded that by selecting proper architecture and hyper-parameters, hybrid models may exceed all other models. Table 4 shows some comparisons of sentiment analysis methods.

4.4. Other approaches

4.4.1. Transformer-based models

Transformer was proposed by Google in 2017 and is one of the most powerful deep-learning model (Parmar et al., 2018). Transformer solves the sequence-to-sequence problem by replacing LSTM with an entire Attention structure, thus achieving better results and reducing computational complexity.

Transformer uses the self-attention mechanism for modelling. With no memory and computing power limitations, the Transformer can theoretically encode infinitely long text. However, because the attention calculation amount is enormous, and the calculation complexity and the sequence length are O(n2), the sequence length increases and the memory and calculation consumption increase rapidly. In practice, due to limited memory and computing power, generally only a certain length can be encoded, such as 512.

In order to improve the long-range encoding capability of

Table 2
Comparative analysis of Lexicon-based approach.

Technique	Advantage	Disadvantage
Dictionary-based approach	There is no need for trained data. Provide positive results for specific domain vocabulary. Quickly access word definitions in the vocabulary.	Opinions have a specific content orientation. Unable to recognize opinion terms with specific content areas that are not yet included in the dictionary.
Corpus-based approach	The ability to recognize distinct content-oriented opinion expressions. The results are better when domains are different.	Performance varies because of the lexicon's wide range. It can't be utilized separately because it is difficult to provide comprehensive texts that can cover every text term.

Transformer and thereby better modeling on long texts, especially document-level corpus, we must optimize the Transformer encoding length.

When the analyzed text length is greater than the max length, the current solution ideas mainly include the following: (1) sequence parallel computing, typical methods such as sequence parallelism (Li et al., 2021a). (2) Optimizing attention mechanisms, typical methods such as Transformer-XL (Dai et al., 2019) and Longformer (Beltagy et al., 2020). (3) Introducing memory mechanisms, typical methods such as Focused Transformer (Beltagy et al., 2020) and Memorizing Transformer (Wu et al., 2022). (4) Introducing a sampling mechanism, typical methods such as Hierarchical transformers (Nawrot et al., 2021) and Dynamic-Pooling Transformer (Nawrot et al., 2022).

A special Transformer model named BERT (Bidirectional Encoder Representations from Transformers) has been developed (Liu et al., 2019). BERT is an NLP Model proposed by Google Research in 2018 (Devlin et al., 2018). It is a core network composed of Transformer Encoder layers, supplemented by word and positional encoding. BERT uses a multi-layer bidirectional transformer encoder instead of Bi-LSTM to encode the syntax and semantics of text comments. The transformer block contains a completely linked layer and a self-attention layer. Reviews are input into the embedding layer of BERT, which uses the contextual information to produce the token-level representation token. The input embedding for a certain input sentence is obtained by location and segment embedding. The token embedding is a unique embedding for each token, whereas the location embedding is the token's position in the review. However, all segmented embedding of sentence tokens point to the specific sentence to which the token belongs. BERT utilizes auxiliary sentiment knowledge by infusing sentiment contextual information into language representation models. The contextual word embedding of this language model includes inter-sentence relationships and understanding the context of the entire comment by preserving the semantic meaning of words in a range of domains. BERT obtains deep bidirectional representations from unmarked reviews. Therefore, by using an additional output layer to improve the pre-trained BERT model, it is possible to generate cutting-edge models suitable for various scenarios (Zhao and Yu, 2021).

BERT is divided into two stages: pre-training and fine-tuning. The model first uses large-scale unlabeled data to obtain the basic model through self-supervised learning. Then, the labeled data of downstream tasks are used to fine-tune the basic model to achieve the adaptation of downstream tasks. Zhou and Srikumar (2021) investigated the BERT family and employed probe strategies to examine how fine-tuning alters the underlying embedding space. There are many improved version of BERT were proposed, such as RoBERTa (Liu et al., 2019), SpanBERT (Joshi et al., 2020), KBERT (Liu et al., 2020), ELECTRA (Clark et al., 2020). These pre-trained models make different improvements to BERT.

Researchers have recently applied BERT to learn how to embed input comments to collect more contextual knowledge. Chen and Huang (2019) enhanced Chinese sentiment analysis using a BERT-based neural network and contextual understanding. Zhao and Yu (2021) improved

Table 3
Comparison of sentiment analysis methods.

Technique	Advantage	Disadvantage
SVM	• Most famous SA algorithm. • Achieve good accuracy for a huge datasets.	• Model fine tuning is quite time-consuming and challenging. • Long training time are required for large datasets.
NB	• Simple to Implement. • Fewer training data are needed. • Less data and training time are needed compared with other methods.	• Assuming that features are mutually independent. • May experience a zero frequency issue. • Limited by imbalanced data categories
ME	• operates using a probabilistic approach. • Less training data are needed.	• The model is domain-oriented and won't work well with other datasets.
DT	• Simple to build. • Less training time. • Training does not require a large datasets.	• Over-fitting is more likely in models. • The domain-oriented model will be built.
KNN	• Non-linear decision boundaries can be constructed. • Without explicit training, data can be continuously added throughout time.	• More datasets and dimensions result in more complex predictions. • All features are given equal weight.
LR	• Simplest models to do classification task.	• Boundaries are linear and unable to handle complex nonlinear problems. • The accuracy of complicated datasets is low.
CNN	• Higher Accuracy. • Faster Training.	• Need a large amount of train datasets and train time. • Pooling layers may result in the feature losing its position or order.
RNN	• Enable to remember long distance relationships between sequential data • High reliability	• Compared to other models, train more slowly. • Costly in terms of computing and complicated.
LSTM	• Better than RNN. • Can capture long term dependencies.	• Extremely complex model. • High training time.
GRU	• More simpler than LSTM. • Faster than LSTM.	• Unable parallel computing.
Transformer	• Self-attention models are used to identify dependencies. • Concentrates only on the sentence's key points.	• Require huge data.
ABSA	• Fine-grained sentiment attribute analysis. • More detailed and specific.	• Lack of large annotated datasets.
Transfer learning	• Improve baseline performance. • Save model development time.	• Model parameters do not converge easily. • Models are prone to over-fitting.
Multi-modal sentiment analysis	• More comprehensive feature representation. • More sources of information allow the model to make better decisions.	• Labeling data is expensive. • Model cross-domain issues.
Multi-task learning	• Reduce the number of models. • Improve data utilization. • Better generalization performance.	• The convergence speed of multiple tasks is inconsistent. • It may be that some sub-tasks work well and some sub-tasks work poorly.

Table 4
The comparison of some sentiment analysis algorithms.

Method	Advantage	Disadvantage
Lexicon-based approach	Simple and easy to understand.	It requires a lot of labor and relies on the accuracy and domain of the emotional dictionary. The accuracy of text sentiment judgment is not high.
Conventional machine learning approach	Ability to classify text emotions based on the sentiment features selection and the emotional classifiers.	Inability to fully utilize the contextual information. To a certain extent, increasing the training set size cannot improve performance.
Deep-learning approach	Can actively learn contextual information. It can extract the semantic information to improve the performance of sentiment analysis. Increasing the training set sizes can improve performance.	A large amount of data support is required. Algorithm training time generally takes a long time. The interpretability of internal structure and theoretical knowledge of deep networks is poor.

the BERT model for ABSA by incorporating external knowledge into contextual embedding as the language representation model.

4.4.2. *Aspect-based sentiment analysis (ABSA)*

ABSA is a significant fine-grained SA problem which has received extensive attention recently. In ABSA problems, the related object expressing emotion is focused on some aspect of an entity (Tan et al., 2020). Generally speaking, the main research directions of ABSA involve identifying sentiment elements at multifaceted levels, namely, aspect terms a, aspect categories c, opinion terms o, and sentiment polarity p (Xu et al., 2020).

Based on whether the expected output is a single sentiment element or a combination of multiple sentiment elements, ABSA tasks can be divided into single ABSA tasks and compound ABSA tasks (Zhang et al., 2022a), e.g., opinion term extraction (OTE) is a single ABSA task which aims to extract opinion term o form given text. However, aspect-opinion pair extraction (AOPE) is a compound ABSA task that aims to extract multiple coupled elements of aspect-opinion from a given text. The overview of the ABSA tasks is depicted in Fig. 5. From this perspective, single ABSA includes aspect term extraction (ATE), opinion term extraction (OTE), aspect category detection (ACD) and aspect sentiment classification (ASC) (Yan et al., 2021; Zhang et al., 2021). ABSA is most commonly used in hotel or product reviews because it allows businesses to detect user opinions on specific aspects (Tran et al., 2019). Zhang et al. (2021) turn aspect-based analytical tasks into text-generation problems. They adopted two modeling paradigms to perform emotion analysis based on generational factors. Chen et al. (2022a) suggested a discrete opinion tree for ABSA that uses grammatical dependencies to capture interactions between aspects and opinion contexts. Table 5 shows an overview of the input and output for each ABSA task, with the example sentence, "Salads were fantastic, but our server was unfriendly.".

Compound ABSA tasks: The target of compound ABSA tasks is extracting and predicting multiple sentiment elements in the sentence.. Compound ABSA tasks include aspect-opinion pair extraction (AOPE), aspect category sentiment analysis (ACSA), end-to-end ABSA (E2E-ABSA), aspect sentiment triplet extraction (ASTE), target aspect sentiment detection (TASD), and aspect-category-opinion-sentiment (ACOS) quadruple extraction. Many studies have been conducted on compound ABSA tasks and we systematically review as follows.

Aspect-Opinion Pair Extraction (AOPE): AOPE task aims to extract the

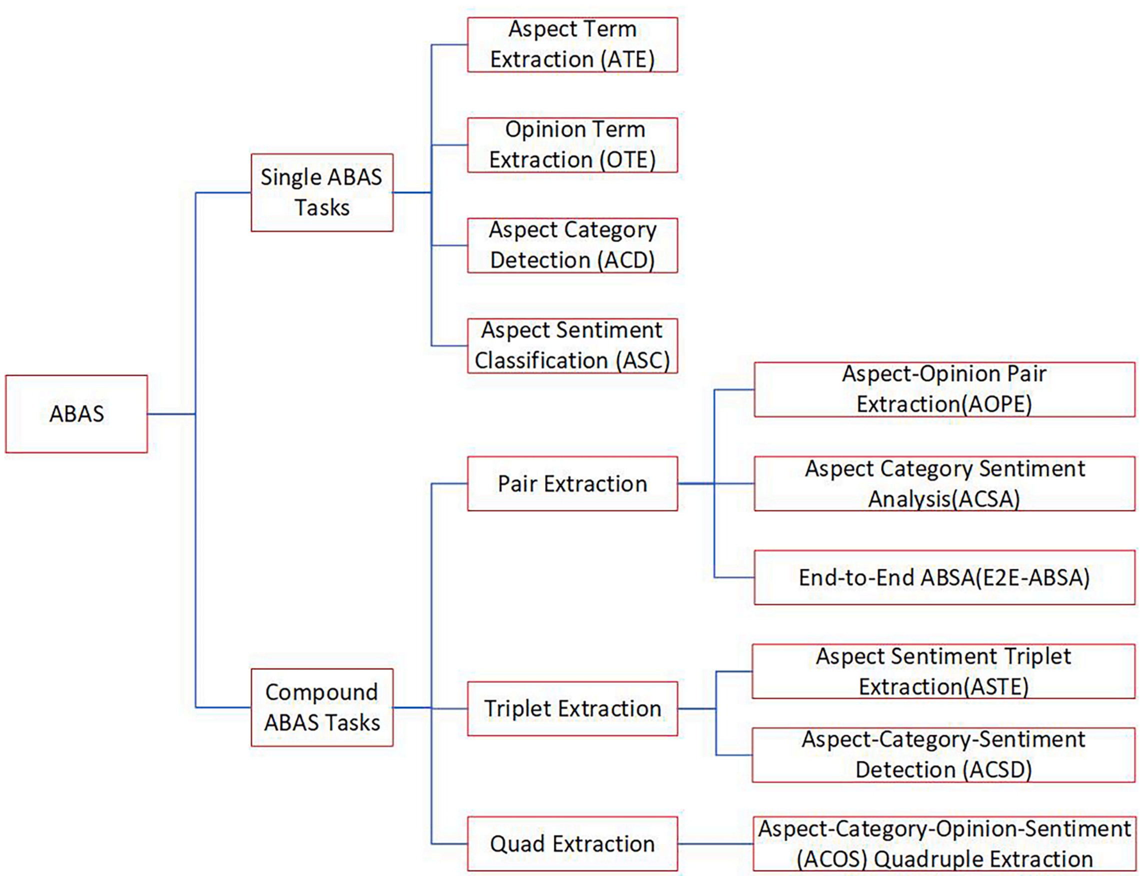

Fig. 5. The ABSA tasks.

Table 5
An Overview of the Input and Output for Each ABSA Task With Examples.

Task	Input	Example Input*	Output	Example Output
Aspect Term Extraction (ATE)	*s*	sentence	*a*	salads,server
Opinion Term Extraction (OTE)	*s*	sentence	*o*	fantastic, unfriendly
Aspect Category Detection (ACD)	*s*	sentence	*c*	food, service
Aspect Sentiment Classification (ASC)	*s,a1* *s,a2*	sentence, salads sentence, server	*p1* *p2*	POS NEG
Aspect-Opinion Pair Extraction (AOPE),	*s*	sentence	*(a,o)*	(salads, fantastic), (server, unfriendly)
Aspect Category Sentiment Analysis (ACSA)	*s*	sentence	*(a,c)*	(salads, food), (server, service)
End-to-End ABSA (E2E-ABSA)	*s*	sentence	*(a,p)*	(salads, POS), (server, NEG)
Aspect Sentiment Triplet Extraction (ASTE)	*s*	sentence	*(a,o,p)*	(salads, fantastic, POS), (server, unfriendly, NEG)
Target Aspect Sentiment Detection (TASD)	*s*	sentence	*(a,c,p)*	(salads, food, POS), (server, service, NEG)
Aspect-Category-Opinion-Sentiment (ACOS) Quadruple Extraction	*s*	sentence	*(a,c,o, p)*	(salads, fantastic, food, POS), (server, unfriendly, service, NEG)

*We assume the example input sentence is: "Salads were fantastic, but our server was unfriendly".

aspect and the corresponding opinion terms in pairs. The AOPE task can adopt the pipeline method to pipe two sub-tasks (ATE and OTE) together to obtain results. Chen et al. (2020) presented a synchronous double-channel RNN to explore the AOPE task. In addition, some researchers construct a joint learning framework from the semantic and grammatical perspectives to extract aspect-opinion pairs (Wang et al., 2022; Zhang et al., 2022b). Another method is to extract the aspect terms (i.e., the ATE task) first and then identify the corresponding opinion terms for each aspect term. Gao et al. (2021) first extract the aspect terms by employing an MRC model, then construct a query to identify the corresponding opinion terms for each aspect term.

Aspect Category Sentiment Analysis (ACSA): The ACSA tasks aim to identify the categories of aspect terms and predict the sentiment polarities. The commonly used method for processing ABSA is the pipeline method: firstly, identify the aspect categories, then predict the sentiment polarities of those categories. Li et al. (2020) proposed a multi-instance multi-label learning network to predict the sentiment of the entity and find the aspect category. Liu et al. (2023) leveraged knowledge enhancement and syntactic data augmentation to enhance ACSA.

End-to-End ABSA (E2E-ABSA): This task aim to extract the aspect terms and predict the corresponding sentiment polarities. E2E-ABSA can be decomposed into two sub-tasks: ATE and ASC (Hu et al., 2019), and the tasks can be conducted sequentially by an intuitive pipeline method. Tian et al. (2021) leverage domain-dependent embedding for End-to-End ABSA. Liang et al. (2021) modeled an interactive architecture based on dependency syntactic knowledge augmented for E2E-ABSA. This model fully exploits the dependency relations and multi-task learning.

Aspect Sentiment Triplet Extraction (ASTE): The ASTE tasks aim to extract aspect terms, corresponding opinion terms and predict sentiment polarities. There are many representative works about ASTE. Peng et al. (2020) decomposed ASTE into two stages. Firstly, the methods extract the aspects, opinion terms and predictions. Secondly, a classifier is leveraged to identify the valid predicted aspects and opinions to construct output triplets. Chen et al. (2022b) tagged aspects, opinions, and emotions, respectively. By identifying the correspondence between target aspects and opinion expressions, sentiment polarities are determined.

Target Aspect Sentiment Detection (TASD): TASD tasks aim to extract aspect terms, identify aspect categories and predict sentiment analysis. To tackle this task, Ke et al. (2023) proposed a contrastive prompts-based data augmentation method that could get high-quality textual representations. To consider the long-distance dependence of triplets and unbalanced label distribution, Ke et al. (2022) combined Prior-BERT with multi-task learning to conduct TASD tasks. Wu et al. (2021) utilized BERT to obtain word embedding, applied BiLSTM to model aspect representations, and utilized graph convolutional network to capture the dependency relationship between aspects and targets.

Aspect-Category-Opinion-Sentiment (ACOS) Quadruple Extraction: This task aims to extract (*a,c, o, p*) quadruples in the given sentence. Compared with other ABSA tasks, the ACOS quadruple extraction task is the most complete ABSA task and extracts all information. The ACOS tasks is important in implicit sentiment analysis. Cai et al. (2021) utilized implicit aspects and opinions to extract all ACOS quadruples. Li et al. (2023) leveraged distance information between aspects and opinions to extract ACOS quadruples in implicit SA. Peper and Wang (2022) utilized contrastive learning to aid the representation of input attributes, then leveraged auto-regressive encoder-decoder models to extract ACOS quadruples.

4.4.3. Transfer learning

Transfer learning is a cutting-edge AI technology that allows a previously learned model to transfer its known information to a new one without any extra training data. Since there is no need to create an algorithm from the start, this method is quite helpful for saving time. This method can transfer knowledge from one field to another and produce good accuracy and results (Celik et al., 2020). This study by Liang et al. (2020) proposes a end-to-end ABSA iterative multi-knowledge transfer network (IMKTN), which fully exploits the interactions in ABSA tasks through knowledge transfer at the token and document layers. Chan et al. (2023) applied sequential transfer learning and made recommendations for future research in the areas of modelling compression, efficient methods for knowledge adaption, neutrality detection, and ambivalence handling.

5. Datasets and evaluation measures

5.1. Datasets

Table 6 provides a variety of datasets that scholars have utilized to evaluate model performance. The most commonly used datasets include SemEval, Stanford Sentiment Treebank (SST), Yelp, and IMDB. Each domain has its version in the SemEval and SST datasets. Yelp datasets include corpora for restaurants, shopping, hotels, travel, etc. IMDB is a commonly used datasets that consist of various movie reviews. ISEAR datasets include one of seven emotions expressed by different respondents in some situations. Texts from social media, review websites, forums, and electronic commerce websites are included in the databases. Many researchers have collected datasets from online social media platforms like Weblog, Twitter, YouTube, and Facebook and classified them into different sentiments by language specialists or scholars. These data are typically unstructured and must be preprocessed before proceeding.

5.2. Evaluation metrics

There are a variety of measures evaluating the efficacy and performance of a method or proposed model. The measurements used in a

Table 6
Sentiment analysis datasets.

Datasets	Data Size	Sentiment	Domain
Stanford sentiment treebank (SST) (Socher et al., 2013)	• SST-1:11,855 reviews • SST-2: 9,613 reviews	• Positive, neutral, negative, very positive, very negative	Movie reviews
Tweets SemEval (Kirange et al., 2014; Nakov et al., 2019; Pontiki et al., 2016; Pontiki et al., 2015; Rosenthal et al., 2019)	• SemEval-2014:7683 reviews • SemEval-2015 • SemEval-2016 • SemEval-2017 • SemEval-2018:7102 tweets	• Positive and negative • Anger, Joy, sad and fear	• Movie reviews • Tweets
Tweets Airline(Rustam et al., 2019)	14,640 samples	Negative, neutral, and positive	Opinions about U.S. airlines
Yelp datasets(Asghar, 2016)	• Yelp 2013:335,018 reviews • Yelp 2014:1,125,457 reviews • Yelp 2018: around 1,000,000 reviews	Positive and negative	Restaurants, shopping, hotels, travel reviews
Stanford large movie review (IMDB) (Maas et al., 2011)	50,000 reviews	Positive and negative	Movie reviews
ISEAR (Poria et al., 2013)	Around 7500 sentences	Positive and negative	Incident reports
Amazon review datasets (Keung et al., 2020)	493MB reviews	Positive and negative	Product reviewson books, electronics, and electric appliance
Cornell Movie Reviews (Maulana et al., 2020)	10,662 reviews	Positive and negative	Movie reviews
OpinRank Datasets (Ganesan, 2010)	• Car review:42230 reviews • Hotel reviews:259,000 reviews	Positive and negative	Car reviews andHotel reviews
CMU-MOSI datasets (Zadeh et al., 2016)	• 2199 opinionated utterances • 93 YuToBe videos	Positive and negative	Multi-modal datasets from different topics
MOUD datasets (Pérez-Rosas et al., 2013)	• 79 videos	Positive and negative	Product reviews
Getty images datasets(You et al., 2016)	• 588,221 labeled images and text	Positive and negative	Product reviews
Twitter datasets(You et al., 2016)	• 220,000 tweets images and text	Positive and negative	Tweet reviews
Twitter image datasets(You et al., 2015)	• 1269 image tweets	Positive and negative	Tweet reviews
MAMS(Jiang et al., 2019)	• ATSA: 17,462 reviews • ACS: 12,592 reviews	Negative, neutral, and positive	Citysearch NewYork datasets

model might have an impact on its success. The main evaluation indicators for sentiment analysis rely on accuracy, F1-score, precision and recall, as briefly described in Table 7.

- TP measures the number of positive samples the classifier successfully identified as positive.
- FP measures the number of negative samples the classifier misclassified as positive.

Table 7
Sentiment analysis evaluation metrics.

Metric	Description	Calculate
Accuracy	Accuracy is the proportion of correctly predicted examples to all examples.	$\frac{TP+TN}{TP+TN+FP+FN}$
Precision	Precision is evaluated as the proportion of correctly categorized positive samples to all positive samples expected.	$\frac{TP}{TP+FP}$
Recall (Sensitivity)	Recall is the percentage of positive samples correctly identified out of all samples that were positive.	$\frac{TP}{TP+FN}$
F1-score (F-measure)	F1-score is calculated by the harmonic mean of precision and recall. F1-score is a number between 0 and 1.	$\frac{2*Precision*Recall}{Precision+Recall}$
Specificity	Specificity is the opposite of sensitivity. The true negative rate quantifies how effectively the negative class was predicted.	$\frac{TN}{TN+FP}$
Mean Square Error (MSE)	MSE is used to measure the divergence between actual label and predict label.	$\frac{\sum_{i=1}^{n}(Actual_i-Predicet_i)^2}{N}$
Mean Absolute Error (MAE)	MAE is utilized to measure the average of the absolute values of the errors between actual value and predict value.	$\frac{\sum_{i=1}^{n}\lvert Actual_i-Predicet_i\rvert}{N}$
Ranking loss	Ranking loss is used to measure the average divergence between actual label and predict label for m sentiment classes and n samples.	$\sum_{i=1}^{n}\frac{\lvert Actual_i-Predicet_i\rvert}{m*n}$
Area Under Curve (AUC)	AUC is defined as the area under the ROC curve and the coordinate axis with M positive samples and N negative samples.	$\frac{\sum I(p_{正样本},p_{负样本})}{M*N}$

- FN measures the number of positive samples the classifier misclassified as negative.
- TN measures the number of negative samples the classifier successfully identified as negative.

In addition to the metrics stated above, for proper interpretation of results, other metrics have been employed to assess SA. Such as Mean Square Error(Verma et al., 2017), Ranking loss (Moghaddam and Ester, 2010), Receiver Operator Characteristic (ROC) (Krupinski, 2017), Area Under the Curve (AUC) (Bowers and Zhou, 2019), Root Mean Square Error (RMSE) (Chai and Draxler, 2014a) and mean absolute error (MAE) (Chai and Draxler, 2014b; Willmott and Matsuura, 2005).

6. Applications of sentiment analysis

6.1. Business domain

SA has many advantages in business analysis. For instance, sentiment analysis may gather client feedback and optimize market strategies to improve their products and services (Birjali et al., 2021). Furthermore, the study may be utilized by consumers to compare products and make better judgments. Therefore, it is not limited to product firms. Wang and Zheng (2016) conducted a study on Chinese internet reviews from various domains to determine what elements may influence sentiment categorization performance. Bose et al. (2020) had collected customer foods reviews on Amazon.com for ten years. The paper used the NRC emotion lexicon and classified reviews into eight emotions (surprise, trust, anticipation, anger, fear, sadness, disgust, and joy).

Sentiment analysis was also used to predict the financial market and forecast stock prices. All market news can be analyzed to predict future stock price trends. There are many ways to collect datasets, such as weibo, blogs, Twitter, news, etc. According to (Xing et al., 2018), news was used to predict whether stock price trends will rise or fall. Trends tended to increase when there was good news, but trends fell when there was terrible news. Rognone et al. (2020) studied the impact of news emotions on the volatility, volume, and returns of conventional currencies and cryptocurrencies like bitcoin.

Due to the booming development of social platforms, leveraging SA technology to analyze social comments and obtain people's sentiment tendencies in the keywords field can help predict future development trends. Baker et al. (2023) researched to detect people's feelings on Twitter during the Russian aggression on Ukraine. The paper first found the datasets about relevant hashtags and then classified the tweet reviews. Jihad et al. (2022) employed machine learning approaches to analyze the opinion of consumers about electric vehicles under the background of rising fuel prices.

6.2. Government intelligence

With the development of online social media, there appeared various comments on politics, life, and social hot issues. Leveraging SA to discover thoughts on social hot problems or policies can help the government monitor potential public reactions. It is beneficial for the government to take corresponding measures or formulate relevant policies. According to (Zavattaro et al., 2015), the research studied tweets from municipal governments in the United States to see how sentiment affects the public's interaction with the government via social media. Sentiment analysis should be conducted in real-time to monitor public sentiment in many cases (El Alaoui et al., 2018; Georgiadou et al., 2020). Falck et al. (2020) calculated the proximity between newspapers and political parties utilizing the Emotional Political Compass (SPC) to examine how political slants in the media affect voters' opinions.

6.3. Healthcare domain

SA has recently been widely employed in the healthcare and medical arena. Healthcare professionals can use this program to collect and analyze data on diseases, adverse drug reactions, epidemics, and patient emotions to provide better medical treatment (Ramírez-Tinoco et al., 2019). In this work (Chintalapudi et al., 2021), applications for text mining that combine sentiment analysis with health information for seafarers are presented. Visualizing patient symptoms is highly beneficial to medical professionals and health organizations. It improves patient knowledge of their issues, keeps track of their medical records, and evaluates their feedback. Baker et al. (2022) leveraged three deep learning methods, LSTM, GRU and CNN, to model colon patient reviews and data. In this research, the future of disease can be predicted by analyzing the Colon cancer datasets.

6.4. Large language models (LLMs)

ChatGPT: It is one of the most advanced large language models developed by OpenAI (Thorp, 2023). It has a high level of fluency and accuracy in comprehending and producing human language (Shen et al., 2023). ChatGPT is the ideal tool for performing sentiment analysis since it understands the complexities and subtleties of human language. When a user inputs text, ChatGPT can learn the semantics of text and the internal relationships between texts. Sentiment analysis combined with ChatGPT offers many benefits to companies and customers. Some of these benefits include real-time feedback, automated analysis, customer focus, cost-effectiveness, and competitive advantage. Sudirjo et al. (2023) provided a study to research the usage of ChatGPT in business customer sentiment analysis. This study found that ChatGPT can aid in understanding and meeting customer needs, preferences and satisfactions and indicated that ChatGPT is beneficial to commercial enterprises. Wang et al. (2023) conducted a evaluation of ChatGPT's ability in understanding opinions and moods contained in the text. This study found that ChatGPT exhibits magnificent ability in sentiment analysis on various evaluation scenarios, and the performance even matches with BERT and state-of-the-art models.

Baidu ERNIE Bot: It leverages SA technologies and this algorithm mainly includes two parts: feature extraction and classifier. Baidu ERNIE Bot employs technologies such as CNN, RNN and Bert to effectively extract emotional elements from the text. Baidu ERNIE Bot's continuous updating and iterative optimization ideas are also crucial for its success. For example, at the end of 2021, Baidu launched a brand new BERT-based sentiment analysis model, namely "Short Text Sentiment Understanding BERT" (SentiBERT) (Yin et al., 2020), which uses the most advanced BERT algorithm in the field of NLP to identify better the emotional information to improve the performance of Baidu ERNIE Bot.

PaLM (Pathways Language Model): It is a language model developed by Google. PaLM-E stands for "Pre-training and Language Model Enhanced", a further improvement based on Google's Bert model. The launch of PaLM-E gives AI the ability to "understand text" and "read images". Compared to the Bert model, the PaLM-E model introduces additional techniques and improvements. One significant improvement is the "bottleneck layer", which can reduce the model's computational burden and improve the training and inference efficiency by using smaller bottleneck layers. In addition, PaLM-E has added a new training objective called "Language Model Enhanced Objective", which can help the model better handle situations such as long sequences and unknown words. Therefore, it can better perform sentiment analysis tasks.

6.5. Scientometric analysis

It aims to analyze the authors' sentiments within scientific citations. This field holds immense potential for leveraging sentiment analysis across various contexts. However, limited researches have been conducted in SA in scientific citations. Yousif et al. (2019) presented a survey of scientific citation sentiment analysis. They proposed some methods and discussed the main challenges. In addition, the authors also introduced relevant fields that have recently attracted great attention, such as citation function classification and citation recommendation.

7. Discussion

In this section, the significant findings, challenges and limitations, and future research directions of SLR are discussed.

7.1. Significant findings

The researchers preferred the general sentiment lexicons over domain-specific sentiment lexicons for SA tasks. The possible reason is that the construction of general sentiment lexicons is more straightforward (Shaukat et al., 2020). In addition, almost all sentiment lexicons used in primary research are universal. One challenge that must be addressed when using sentiment lexicons is that they must be constantly upgraded with the development of society. In addition, lexicon-based methods must handle negative phrases, spelling errors, synonyms, and slang, as they require precise matching between textual and dictionary terms.

Most researchers chose conventional machine learning and deep learning methods for SA tasks to overcome the limitations of previous lexicon-based techniques. SVM is the most frequently used machine learning in SA tasks and obtains high accuracy on many public datasets. The possible reason may be SVM's ability to handle high-dimensional text features. Nevertheless, the drawback of SVM is that the training time is longer as the size of training datasets increases.

Bi-LSTM is the most frequently used deep learning architecture for SA tasks in primary research. One possible reason is that Bi LSTM has the ability to capture long-term dependencies, especially when dealing with continuous data such as text. Another reason may be the effective training time of Bi-LSTM on large datasets.

Transformer-based pre-trained language models such as BERT has been demonstrated to perform well and are widely used in SA tasks (Zhou and Srikumar, 2021). The possible cause is that the BERT is pre-trained on large-scale text data and then fine-tuned on specific tasks, significantly reducing training time and data requirements. Fine-tuning BERT for special tasks may not require many additional datasets. Additionally, BERT's bidirectional Transformer encoder nature enables it to distinguish words with similar spellings but different meanings well. It is more context-aware, which can effectively capture the dependency relationships between contexts, thereby improving the accuracy of SA tasks. (Devlin et al., 2018). Since the SA tasks are different from other NLP tasks, they require more semantic tokens and contextual information. Therefore, paying attention to essential tokens that affect the sentiment of the entire sentence context is necessary, and selecting these tokens is challenging. When suitable adjustments are made to the output layer of the BERT model, the semantics can be inferred from the text, and contextual information can be learned. Sometimes, research work is prohibited due to high computational requirements. To overcome these issues, the experiment was conducted by discarding different parts of the original architecture, and even in the case of minimizing the model, removing the alternating layers of BERT can achieve better results. Applications in fields such as finance and medicine that lack domain data have significantly benefited from BERT pre-trained language models. Similarly, many Chinese, Arabic, German, French and Russian languages utilized the BERT model to conduct SA and achieved perfect performance.

ABSA is a fine-grained SA research that has achieved significant performance improvements in recent years. However, there still are several issues. If the usage domain of the ABSA model is different from the field of training datasets, mainstream methods reveal the domain adaptation problem of performance degradation. Nevertheless, the vast sample corpus for each task domain must train the correct model to capture appropriate contextual aspects. Almost half of the ABSA studies used public benchmark datasets from SemEval in two domains, namely laptops and restaurants. However, the most effective ABSA model uses independently labeled data in each domain. Since the same sentiment words may have different meanings in different fields (Chauhan et al., 2023). Therefore, ABSA models trained in specific fields often perform unsatisfactorily on other domain datasets. In addition, baseline technologies cannot associate adequate multiple-aspect words extracted from different comment sentences, thus extracting invalid aspect terms. Invalid aspect extraction can decrease aspect extraction performance (Poria et al., 2016). Collecting correlation information between the sentence and considering context can help determine the exact meaning of the aspect terms and opinion terms. In addition, BERT representation and models based on supervised hierarchical attention may be helpful for aspect terms extraction. It is possible to add sentence co-reference resolution steps as preprocessing before performing ABSA.

7.2. Challenges and limitations

According to the above study's findings, it is necessary to establish robust datasets in many languages. The datasets should be well annotated and finely graded. The compilation and analysis of datasets should comply with ethical standards and be widely available in the public domain for better research. The co-reference resolution problem needs to be focused on, and detecting hidden emotions, irony, and sarcasm remains an open research question in SA. Feature extraction in SA faces several issues, including context dependency, high dimensionality, redundancy and slang words. In addition, open research topics in SA include handling multilingual data, cross-domain accuracy improvement, cross-datasets SA, and implicit SA using contextual backgrounds. Especially, two hot research topics, namely ABSA and BERT, remain many challenges and limitations. Some of the challenges are listed below.

Sarcasm and Ridicule: The problem of identifying sarcasm and ridicule contained in text has long been a complex problem for researchers in the field of SA. Because these feelings are not clearly expressed in the text, they are called implicit emotions. People perceive these emotions through two cues: voice communication and textual context. Multi-modal SA can detect nonverbal cues and context for implicit emotion identification. Multi-modal SA combines two or three input types from text, audio, and images to improve the accuracy of SA. A possible future research approach is to combine multi-modal cues and context to identify implicit sentiment expressions in text.

Context Dependency: Some words are inherently objective but may rely on subjectivity when used in a particular setting or context and thus be emotionally charged (Denecke and Deng, 2015). For instance, the word "long" is an adjective that does not contain sentiment. However, it may be positive in some instances (such as "long battery") or negative in others (such as "long queue"). Frequently, a word's context depends on terms outside the range of adjacent words because they are located far away in the phrase. Word vectors and parse tree models work well together to discover large sub-structures.

High Dimensionality: It refers to a large number of feature sets that reduce performance due to computational problems, requiring the correct feature selection methods (Wilson et al., 2005).

Slang and abbreviations: Internet users frequently utilize slang and abbreviations in their messages (Wu et al., 2018). According to Wikipedia (2014), slang is a language made up of non-standard words and expressions like GR8, SMH, YYDS, and XOXO. The main reason people use slang is because it is convenient, frequently easy to understand, and can save time. In WeChat, Twitter, and Facebook messaging, a significant number of slang words with either good or negative sentiments are employed (Asghar et al., 2014). Detecting, interpreting, and identifying Slang's polarity is now crucial for figuring out the review's semantic orientation (SO). This paper described a methodology for detecting and scoring Internet slang (DSIS) by combining SentiWordNet with additional lexical resources (Kundi et al., 2014).

Cross Domain: One major challenge faced by SA is its inability to perform well in other domains due to insufficient labeled data and domain knowledge. The training datasets of current sentiment analysis methods are always domain-dependent. This model has poor generalization and unsatisfactory performance in new fields. Moreover, using datasets from every field to train models is also impractical. Domain adaptation can solve the above problems by learning the features of invisible domains. Recently, researchers have tended to use word embedding to solve domain adaptation problems, such as BERT and Glove. Because these word embedding are trained on open-domain datasets, they contain many domain-invariant information. Therefore, the key to the current Domain Adaptation problem lies in the breakthrough of open domain corpus, which also brings about the problem of interpretability.

Multilingual data: Although most SA methods currently focus on English corpus, with the evolution of global social networks, comments are presented in a multilingual state. For example, the comment, "Oh, my god! 这个演唱会太cool了吧! Chiar îmi place." This sentence contains Chinese, English and Roman orthography. This code-mixed data poses new challenges for sentiment analysis. The current possible solution is to develop a unique language model applied to code-mixed data. In addition, machine translation can also be used as a solution for multilingual sentiment analysis. Saadany and Orasan (2020) pointed out that integrating sentiment information in the coding phase of machine translation can solve the problem of not preserving negation, antonym and habitual expression in cross-lingual translation.

Major challenges of ABSA: Most existing ABSA datasets are sourced from SemEval. One potential reason why ABSA performs poorly in multi-lingual environments is that, apart from SemEval serial datasets, there needs to be more review datasets available in other languages, such as Chinese and Arabic. However, relatively small amounts of data make it challenging to train high-performance models, such as PLM-based models with millions of parameters. Therefore, more challenging datasets from multi-domain and multi-lingual are required.

Furthermore, integrating contextual information to the aspects of reviews remains a significant challenge. Due to the lack of context and domain information, extracting the most relevant aspect phrases for a specific domain is difficult. Due to the current technology's inability to extract some vital aspect terms, capturing long-term dependencies between noun phrases is impossible. Although deep learning methods can locate relevant n-grams, current methods are unable to capture sentence relationships for accurate extraction of multiple aspect terms, further hindering the extraction of aspect-opinion categories. For a long review text, it is difficult to extract effective and multi-aspect terms based on establishing connections between different terms and contexts of different phrases. In addition, the evidence provided by Lo et al. (2017) suggests that multilingual aspect extractions face significant challenges, such as word ambiguity, language-dependent structures of sentences, and translation problems. The efficiency of models varies greatly across languages, and methods that perform well in one language may not well in other languages. Therefore, significant effort is required to handle ABSA across language tasks. In summary, ABSA is still in its infancy. With the deepening of research, people will gradually tackle those issues.

Challenges of BERT: Currently, BERT is a Transformer-based model with better performance in SA tasks. However, it requires a large amount of computing resources and time to train and use BERT, because BERT is a relatively large model with hundreds of millions of parameters. Although the BERT model has been pre-trained on large-scale corpora, it may suffer from insufficient learning in certain tasks, leading to performance degradation. And the ability to learn negative compound words is poor. In addition, the pre-training datasets of the BERT model mainly comes from large English corpora and has relatively little support for other languages, which may affect its performance on other languages. Considering the solutions provided by various BERT variants to address BERT's related weaknesses, this article suggests using various integration techniques to integrate these BERT variants to provide universal and diverse outputs. In view with there still remains an enormous amount of unlabeled data in the field of SA, a robust semi-supervised and self-supervised learning architecture could be in-depth explored. In addition, the use of the GPT model for this task is highly encouraged. This is because GPT possesses lexical robustness and is able to produce better performance in related tasks.

7.3. Suggestions and future research directions

However, many gaps still need to be addressed in SA. This research provided some suggestions as a reference for researchers from various perspectives:

1) Datasets: Automatic annotation, Domain adaptation, Multi-emotion classes, Polarity fuzziness, and Unified dictionary or datasets.
2) ABSA: Questing for larger and more challenging datasets, Cross-domain-transfer learning, Unified model for multiple tasks.

3) Feature extraction: Negation detection, Labeled with sarcasm, Emoticons handling, Stop words automatic detection.

4) Sentiment analysis approaches: Generic cross-domain sentiment classifiers, Pragmatic analysis and Hybrid approach for Multilingual data.

Given the discussed challenges and limitations of the SA tasks, we can explore different future research directions for SA:

- Due to ABSA domain dependency and to evaluate the universality of the model on cross-domain data, it is necessary to construct multi-domain and multi-lingual benchmark datasets.
- Cross-task transfer has been proven effective in transferring knowledge learned from low-level ABSA tasks to high-level ABSA tasks (Zhang et al., 2022a). Therefore, it is worthwhile exploring cross-task transfer learning in different types of ABSA tasks.
- Due to the good performance of combining sentiment lexicons with other machine learning methods on SA (Beigi and Moattar, 2021), it is recommended to develop self-supervised techniques for automatic or semi-automatic construction of domain-dependent sentiment lexicons.
- Deeply applying SA in under-explored areas such as healthcare, education and social media, or unexplored areas such as aviation, tourism, and transportation services.
- Conducting additional research to compare various large language models. For example, research can include comparisons of GPT, Baidu ERNIE Bot and PaLM at different granularity levels for all SA tasks.
- Creating customized dialectical grammar analyzers to resolve ambiguity in some words and consider compound words, such as negation and enhancement words.
- Considering that some implicit text expressions are challenging to recognize emotions, it is worth researching solutions to the issue, such as data augmentation (Liu et al., 2023), knowledge enhancement (Zhao et al., 2022) and context awareness (Xu et al., 2022a).

8. Conclusion

This article provides a systematic literature review of text sentiment analysis. Firstly, an introduction was given to the levels of sentiment analysis, including document-level, sentence-level and aspect-level. Secondly, this paper categorized various SA techniques and compared their advantages and drawbacks. Then, we described the applications of SA such as it can be employed in business, government intelligence, healthcare domain and recommendation systems. Besides, the state-of-the-art SA large language models are introduced, including ChatGPT, ERNIE Bot and PaLM. Finally, we elaborated on the current challenges sentiment analysis faces in detail, demonstrating that sentiment analysis still has many unresolved issues to be studied.

CRediT authorship contribution statement

Yanying Mao: Writing – original draft, Software, Methodology, Investigation, Formal analysis. **Qun Liu:** Writing – review & editing, Supervision, Resources, Methodology, Formal analysis. **Yu Zhang:** Writing – review & editing, Supervision, Funding acquisition.

Declaration of competing 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 work was supported by the National Natural Science Foundation of China (Grant No. 72204033), the Science and Technology Research Project of Chongqing Education Commission (Grant No. KJQN202303120), the Humanities and Social Science project of Ministry of Education of China (Grant No. 21YJC630169), the China Postdoctoral Science Foundation (Grant No. 2022M711457) and the Natural Science Foundation of Chongqing (Grant No. cstc2021jcyj-msxmX1010).

Declarations

Ethical approval Not applicable.

Consent to participate Not applicable.

Consent for publication The authors concur the publication by Journal of King Saud University-Computer and Information Sciences.

References

Abdelgwad, M.M., Soliman, T.H.A., Taloba, A.I., Farghaly, M.F., 2022. Arabic aspect based sentiment analysis using bidirectional GRU based models. Journal of King Saud University-Computer and Information Sciences 34, 6652–6662.
Agarwal, B., Mittal, N., Agarwal, B., Mittal, N., 2016. Machine learning approach for sentiment analysis. Prominent Feature Extraction for Sentiment Analysis 21–45.
Ahmad, M., Aftab, S., Ali, I., 2017. Sentiment analysis of tweets using svm. Int. J. Comput. Appl 177, 25–29.
Alshuwaier, F., Areshey, A., Poon, J., 2022. Applications and enhancement of document-based sentiment analysis in deep learning methods: Systematic literature review. Intelligent Systems with Applications 15, 200090.
Altaf, A., Anwar, M.W., Jamal, M.H., Bajwa, U.I., 2023. Exploiting Linguistic Features for Effective Sentence-Level Sentiment Analysis in Urdu Language. Multimed. Tools Appl. 1–27.
Alyami, S., Alhothali, A., Jamal, A., 2022. Systematic literature review of Arabic aspect-based sentiment analysis. Journal of King Saud University-Computer and Information Sciences 34, 6524–6551.
Appel, O., Chiclana, F., Carter, J., Fujita, H., 2016. A hybrid approach to the sentiment analysis problem at the sentence level. Knowl.-Based Syst. 108, 110–124.
Asghar, M.Z., Khan, A., Ahmad, S., Kundi, F.M., 2014. A review of feature extraction in sentiment analysis. Journal of Basic and Applied Scientific Research 4, 181–186.
Asghar, N., 2016. Yelp dataset challenge: Review rating prediction. arXiv preprint arXiv: 1605.05362.
Baker, M.R., Mohammed, E.Z., Jihad, K.H., 2022. Prediction of Colon Cancer Related Tweets Using Deep Learning Models, International Conference on Intelligent Systems Design and Applications. Springer, pp. 522-532.
Baker, M.R., Taher, Y.N., Jihad, K.H., 2023. PREDICTION OF PEOPLE SENTIMENTS ON TWITTER USING MACHINE LEARNING CLASSIFIERS DURING RUSSIAN AGGRESSION IN UKRAINE. Jordanian Journal of Computers & Information Technology 9.
Behdenna, S., Barigou, F., Belalem, G., 2016. Sentiment analysis at document level, Smart Trends in Information Technology and Computer Communications: First International Conference, SmartCom 2016, Jaipur, India, August 6–7, 2016, Revised Selected Papers 1. Springer 159–168.
Behdenna, S., Barigou, F., Belalem, G., 2018. Document level sentiment analysis: a survey. EAI Endorsed Transactions on Context-Aware Systems and Applications 4, e2–e.
Beigi, O.M., Moattar, M.H., 2021. Automatic construction of domain-specific sentiment lexicon for unsupervised domain adaptation and sentiment classification. Knowl.-Based Syst. 213, 106423.
Beltagy, I., Peters, M.E., Cohan, A., 2020. Longformer: The long-document transformer. arXiv preprint arXiv:2004.05150.
Birjali, M., Kasri, M., Beni-Hssane, A., 2021. A comprehensive survey on sentiment analysis: Approaches, challenges and trends. Knowl.-Based Syst. 226, 107134.
Boiy, E., Moens, M.-F., 2009. A machine learning approach to sentiment analysis in multilingual Web texts. Inf. Retr. 12, 526–558.
Bose, R., Dey, R.K., Roy, S., Sarddar, D., 2020. Sentiment analysis on online product reviews, Information and Communication Technology for Sustainable Development: Proceedings of ICT4SD 2018. Springer, pp. 559-569.
Bowers, A.J., Zhou, X., 2019. Receiver operating characteristic (ROC) area under the curve (AUC): A diagnostic measure for evaluating the accuracy of predictors of education outcomes. Journal of Education for Students Placed at Risk (JESPAR) 24, 20–46.
Bu, K., Liu, Y., Ju, X., 2023. Efficient Utilization of Pre-trained Models: A Review of Sentiment Analysis via Prompt Learning. Knowl.-Based Syst. 111148.
Cai, H., Xia, R., Yu, J., 2021. Aspect-category-opinion-sentiment quadruple extraction with implicit aspects and opinions. In: Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (volume 1: Long Papers), pp. 340–350.
Celik, Y., Talo, M., Yildirim, O., Karabatak, M., Acharya, U.R., 2020. Automated invasive ductal carcinoma detection based using deep transfer learning with whole-slide images. Pattern Recogn. Lett. 133, 232–239.
Chai, T., Draxler, R.R., 2014a. Root mean square error (RMSE) or mean absolute error (MAE). Geosci. Model Dev. Discuss. 7, 1525–1534.
Chai, T., Draxler, R.R., 2014b. Root mean square error (RMSE) or mean absolute error (MAE)?–Arguments against avoiding RMSE in the literature. Geosci. Model Dev. 7, 1247–1250.
Chan, J.-Y.-L., Bea, K.T., Leow, S.M.H., Phoong, S.W., Cheng, W.K., 2023. State of the art: a review of sentiment analysis based on sequential transfer learning. Artif. Intell. Rev. 56, 749–780.
Chang, J.-R., Liang, H.-Y., Chen, L.-S., Chang, C.-W., 2020. Novel feature selection approaches for improving the performance of sentiment classification. J. Ambient Intell. Hum. Comput. 1–14.
Chaturvedi, I., Cambria, E., Welsch, R.E., Herrera, F., 2018. Distinguishing between facts and opinions for sentiment analysis: Survey and challenges. Information Fusion 44, 65–77.
Chauhan, G.S., Nahta, R., Meena, Y.K., Gopalani, D., 2023. Aspect based sentiment analysis using deep learning approaches: A survey. Computer Science Review 49, 100576.
Chen, Y., 2015. Convolutional neural network for sentence classification. University of Waterloo.
Chen, F., Huang, Y., 2019. Knowledge-enhanced neural networks for sentiment analysis of Chinese reviews. Neurocomputing 368, 51–58.
Chen, S., Liu, J., Wang, Y., Zhang, W., Chi, Z., 2020. Synchronous double-channel recurrent network for aspect-opinion pair extraction. In: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, pp. 6515–6524.
Chen, C., Teng, Z., Wang, Z., Zhang, Y., 2022a. Discrete opinion tree induction for aspect-based sentiment analysis. In: Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (volume 1: Long Papers), pp. 2051–2064.
Chen, T., Xu, R., He, Y., Wang, X., 2017. Improving sentiment analysis via sentence type classification using BiLSTM-CRF and CNN. Expert Syst. Appl. 72, 221–230.
Chen, F., Yang, Z., Huang, Y., 2022b. A multi-task learning framework for end-to-end aspect sentiment triplet extraction. Neurocomputing 479, 12–21.
Chintalapudi, N., Battineni, G., Di Canio, M., Sagaro, G.G., Amenta, F., 2021. Text mining with sentiment analysis on seafarers' medical documents. International Journal of Information Management Data Insights 1, 100005.
Clark, K., Luong, M.-T., Le, Q.V., Manning, C.D., 2020. Electra: Pre-training text encoders as discriminators rather than generators. arXiv preprint arXiv:2003.10555.
Cortes, C., Vapnik, V., 1995. Support-Vector Networks. Machine Learning 20, 273–297.
Daeli, N.O.F., Adiwijaya, A., 2020. Sentiment analysis on movie reviews using Information gain and K-nearest neighbor. Journal of Data Science and Its Applications 3, 1–7.
Dai, Z., Yang, Z., Yang, Y., Carbonell, J., Le, Q.V., Salakhutdinov, R., 2019. Transformer-xl: Attentive language models beyond a fixed-length context. arXiv preprint arXiv: 1901.02860.
Das, R., Singh, T.D., 2023. Multimodal sentiment analysis: A survey of methods, trends and challenges. ACM Comput. Surv.
Denecke, K., Deng, Y., 2015. Sentiment analysis in medical settings: New opportunities and challenges. Artif. Intell. Med. 64, 17–27.
Devlin, J., Chang, M.-W., Lee, K., Toutanova, K., 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv: 1810.04805.
Do, H.H., Prasad, P.W., Maag, A., Alsadoon, A., 2019. Deep learning for aspect-based sentiment analysis: a comparative review. Expert Syst. Appl. 118, 272–299.
El Alaoui, I., Gahi, Y., Messoussi, R., Chaabi, Y., Todoskoff, A., Kobi, A., 2018. A novel adaptable approach for sentiment analysis on big social data. Journal of Big Data 5, 1–18.
Eskandari, F., Shayestehmanesh, H., Hashemi, S., 2015. Predicting best answer using sentiment analysis in community question answering systems, 2015 Signal Processing and Intelligent Systems Conference (SPIS). IEEE 53–57.
Falck, F., Marstaller, J., Stoehr, N., Maucher, S., Ren, J., Thalhammer, A., Rettinger, A., Studer, R., 2020. Measuring proximity between newspapers and political parties: the sentiment political compass. Policy Internet 12, 367–399.
Gandhi, A., Adhvaryu, K., Poria, S., Cambria, E., Hussain, A., 2023. Multimodal sentiment analysis: A systematic review of history, datasets, multimodal fusion methods, applications, challenges and future directions. Information Fusion 91, 424–444.
Ganesan, K., 2010. Opinrank review dataset.
Gao, L., Wang, Y., Liu, T., Wang, J., Zhang, L., Liao, J., 2021. Question-driven span labeling model for aspect–opinion pair extraction. In: Proceedings of the AAAI Conference on Artificial Intelligence, pp. 12875–12883.
Georgiadou, E., Angelopoulos, S., Drake, H., 2020. Big data analytics and international negotiations: Sentiment analysis of Brexit negotiating outcomes. Int. J. Inf. Manag. 51, 102048.
Hochreiter, S., Schmidhuber, J., 1997. Long short-term memory. Neural Comput. 9, 1735–1780.
Hu, M., Peng, Y., Huang, Z., Li, D., Lv, Y., 2019. Open-domain targeted sentiment analysis via span-based extraction and classification. arXiv preprint arXiv: 1906.03820.
Huang, H., Asemi, A., Mustafa, M.B., 2023. Sentiment analysis in e-commerce platforms: A review of current techniques and future directions. IEEE Access.
Huang, B., Guo, R., Zhu, Y., Fang, Z., Zeng, G., Liu, J., Wang, Y., Fujita, H., Shi, Z., 2022a. Aspect-level sentiment analysis with aspect-specific context position information. Knowl.-Based Syst. 243, 108473.
Huang, W., Lin, M., Wang, Y., 2022b. Sentiment analysis of Chinese e-commerce product reviews using ERNIE word embedding and attention mechanism. Appl. Sci. 12, 7182.
Jain, P.K., Pamula, R., Srivastava, G., 2021. A systematic literature review on machine learning applications for consumer sentiment analysis using online reviews. Computer Science Review 41, 100413.
Jiang, Q., Chen, L., Xu, R., Ao, X., Yang, M., 2019. A challenge dataset and effective models for aspect-based sentiment analysis. In: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), pp. 6280–6285.
Jihad, K.H., Baker, M.R., Farhat, M., Frikha, M., 2022. Machine Learning-Based Social Media Text Analysis: Impact of the Rising Fuel Prices on Electric Vehicles, International Conference on Hybrid Intelligent Systems. Springer, pp. 625-635.
Joshi, M., Chen, D., Liu, Y., Weld, D.S., Zettlemoyer, L., Levy, O., 2020. Spanbert: Improving pre-training by representing and predicting spans. Transactions of the Association for Computational Linguistics 8, 64–77.
Kang, H., Yoo, S.J., Han, D., 2012. Senti-lexicon and improved Naïve Bayes algorithms for sentiment analysis of restaurant reviews. Expert Syst. Appl. 39, 6000–6010.
Ke, C., Xiong, Q., Wu, C., Liao, Z., Yi, H., 2022. Prior-bert and multi-task learning for target-aspect-sentiment joint detection. In: ICASSP 2022–2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, pp. 7817–7821.
Ke, C., Xiong, Q., Wu, C., Yi, H., Gao, M., Chen, J., 2023. SimCPD: a simple framework for contrastive prompts of target-aspect-sentiment joint detection. Neural Comput. & Applic. 1–16.

Keung, P., Lu, Y., Szarvas, G., Smith, N.A., 2020. The multilingual amazon reviews corpus. arXiv preprint arXiv:2010.02573.
Khoshnevisan, B., Rafiee, S., Iqbal, J., Shamshirband, S., Omid, M., BADRUL, A.N., ABDUL, W.A., 2015. A comparative study between artificial neural networks and adaptive neuro-fuzzy inference systems for modeling energy consumption in greenhouse tomato production: A case study in isfahan province.
Kingsford, C., Salzberg, S.L., 2008. What are decision trees? Nat. Biotechnol. 26, 1011–1013.
Kirange, D., Deshmukh, R.R., Kirange, M., 2014. Aspect based sentiment analysis semeval-2014 task 4. Asian Journal of Computer Science and Information Technology (AJCSIT) 4.
Kleinbaum, D.G., Dietz, K., Gail, M., Klein, M., Klein, M., 2002. Logistic regression. Springer.
Krupinski, E.A., 2017. Receiver Operating Characteristic (ROC) Analysis. Frontline Learn. Res. 5, 41–52.
Kundi, F.M., Ahmad, S., Khan, A., Asghar, M.Z., 2014. Detection and scoring of internet slangs for sentiment analysis using SentiWordNet. Life Science Journal 11, 66–72.
Li, Y., Yin, C., Zhong, S.-h., Pan, X., 2020. Multi-instance multi-label learning networks for aspect-category sentiment analysis. arXiv preprint arXiv:2010.02656.
Li, S., Xue, F., Baranwal, C., Li, Y., You, Y., 2021a. Sequence parallelism: Long sequence training from system perspective. arXiv preprint arXiv:2105.13120.
Li, J., Li, X., Du, Y., Fan, Y., Chen, X., Huang, D., 2023. An Aspect-Category-Opinion-Sentiment Quadruple Extraction with Distance Information for Implicit Sentiment Analysis. Information Technology and Control 52, 445–456.
Liang, Y., Meng, F., Zhang, J., Chen, Y., Xu, J., Zhou, J., 2020. An iterative multi-knowledge transfer network for aspect-based sentiment analysis. arXiv preprint arXiv:2004.01935.
Liang, Y., Meng, F., Zhang, J., Chen, Y., Xu, J., Zhou, J., 2021. A dependency syntactic knowledge augmented interactive architecture for end-to-end aspect-based sentiment analysis. Neurocomputing 454, 291–302.
Liu, B., 2012. Sentiment analysis and opinion mining. Synthesis Lectures on Human Language Technologies 5, 1–167.
Liu, P., Qiu, X., Huang, X., 2016. Recurrent neural network for text classification with multi-task learning. arXiv preprint arXiv:1605.05101.
Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., Stoyanov, V., 2019. Roberta: A robustly optimized bert pretraining approach. arXiv preprint arXiv:1907.11692.
Liu, B., Lin, T., Li, M., 2023. Enhancing aspect-category sentiment analysis via syntactic data augmentation and knowledge enhancement. Knowl.-Based Syst. 264, 110339.
Liu, B., Zhang, L., 2012. A survey of opinion mining and sentiment analysis. Mining Text Data. Springer 415–463.
Liu, W., Zhou, P., Zhao, Z., Wang, Z., Ju, Q., Deng, H., Wang, P., 2020. K-bert: Enabling language representation with knowledge graph. In: Proceedings of the AAAI Conference on Artificial Intelligence, pp. 2901–2908.
Lo, S.L., Cambria, E., Chiong, R., Cornforth, D., 2017. Multilingual sentiment analysis: from formal to informal and scarce resource languages. Artif. Intell. Rev. 48, 499–527.
Maas, A., Daly, R.E., Pham, P.T., Huang, D., Ng, A.Y., Potts, C., 2011. Learning word vectors for sentiment analysis. In: Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, pp. 142–150.
Madhoushi, Z., Hamdan, A.R., Zainudin, S., 2015. Sentiment analysis techniques in recent works, 2015 science and information conference (SAI). IEEE 288–291.
Mao, Y., Zhang, Y., Jiao, L., Zhang, H., 2022. Document-level sentiment analysis using attention-based bi-directional long short-term memory network and two-dimensional convolutional neural network. Electronics 11, 1906.
Marong, M., Batcha, N.K., Mafas, R., 2020. Sentiment Analysis in E-Commerce: A Review on The Techniques and Algorithms. Journal of Applied Technology and Innovation (e-ISSN 2600–7304) 4, 6.
Matrane, Y., Benabbou, F., Sael, N., 2023. A systematic literature review of Arabic dialect sentiment analysis. Journal of King Saud University-Computer and Information Sciences, 101570.
Maulana, R., Rahayuningsih, P.A., Irmayani, W., Saputra, D., Jayanti, W.E., 2020. Improved accuracy of sentiment analysis movie review using support vector machine based information gain. Journal of Physics: Conference Series. IOP Publishing, 012060.
Mercha, E.M., Benbrahim, H., 2023. Machine learning and deep learning for sentiment analysis across languages: A survey. Neurocomputing 531, 195–216.
Mitra, A., 2020. Sentiment analysis using machine learning approaches (Lexicon based on movie review dataset). Journal of Ubiquitous Computing and Communication Technologies (UCCT) 2, 145–152.
Moghaddam, S., Ester, M., 2010. Opinion digger: an unsupervised opinion miner from unstructured product reviews. In: Proceedings of the 19th ACM International Conference on Information and Knowledge Management, pp. 1825–1828.
Moraes, R., Valiati, J.F., Neto, W.P.G., 2013. Document-level sentiment classification: An empirical comparison between SVM and ANN. Expert Syst. Appl. 40, 621–633.
Mubarok, M.S., Adiwijaya, A., M.d.,, 2017. Aspect-based sentiment analysis to review products using Naïve Bayes. AIP Publishing LLC, AIP conference proceedings, p. 020060.
Myles, A.J., Feudale, R.N., Liu, Y., Woody, N.A., Brown, S.D., 2004. An introduction to decision tree modeling. Journal of Chemometrics: A Journal of the Chemometrics Society 18, 275–285.
Nakov, P., Ritter, A., Rosenthal, S., Sebastiani, F., Stoyanov, V., 2019. SemEval-2016 task 4: Sentiment analysis in Twitter. arXiv preprint arXiv:1912.01973.
Nawrot, P., Tworkowski, S., Tyrolski, M., Kaiser, Ł., Wu, Y., Szegedy, C., Michalewski, H., 2021. Hierarchical transformers are more efficient language models. arXiv preprint arXiv:2110.13711.
Nawrot, P., Chorowski, J., Łańcucki, A., Ponti, E.M., 2022. Efficient transformers with dynamic token pooling. arXiv preprint arXiv:2211.09761.
Nigam, K., Lafferty, J., McCallum, A., 1999. Using maximum entropy for text classification, IJCAI-99 workshop on machine learning for information filtering. Stockholom, Sweden, pp. 61–67.
Osorio Angel, S., Negrón, P.P.A., Espinoza-Valdez, A., 2021. Systematic literature review of sentiment analysis in the Spanish language. Data Technologies and Applications 55, 461–479.
Oueslati, O., Cambria, E., HajHmida, M.B., Ounelli, H., 2020. A review of sentiment analysis research in Arabic language. Futur. Gener. Comput. Syst. 112, 408–430.
Parmar, N., Vaswani, A., Uszkoreit, J., Kaiser, L., Shazeer, N., Ku, A., Tran, D., 2018. Image transformer. International Conference on Machine Learning. PMLR 4055–4064.
Peng, H., Cambria, E., Hussain, A., 2017. A review of sentiment analysis research in Chinese language. Cogn. Comput. 9, 423–435.
Peng, H., Xu, L., Bing, L., Huang, F., Lu, W., Si, L., 2020. Knowing what, how and why: A near complete solution for aspect-based sentiment analysis. In: Proceedings of the AAAI Conference on Artificial Intelligence, pp. 8600–8607.
Peper, J.J., Wang, L., 2022. Generative aspect-based sentiment analysis with contrastive learning and expressive structure. arXiv preprint arXiv:2211.07743.
Pereira, D.A., 2021. A survey of sentiment analysis in the Portuguese language. Artif. Intell. Rev. 54, 1087–1115.
Pérez-Rosas, V., Mihalcea, R., Morency, L.-P., 2013. Utterance-level multimodal sentiment analysis, Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pp. 973-982.
Phan, H.T., Nguyen, N.T., Hwang, D., 2023. Aspect-level sentiment analysis: A survey of graph convolutional network methods. Information Fusion 91, 149–172.
Pontiki, M., Galanis, D., Papageorgiou, H., Manandhar, S., Androutsopoulos, I., 2015. Semeval-2015 task 12: Aspect based sentiment analysis, Proceedings of the 9th international workshop on semantic evaluation (SemEval 2015), pp. 486-495.
Pontiki, M., Galanis, D., Papageorgiou, H., Androutsopoulos, I., Manandhar, S., AL-Smadi, M., Al-Ayyoub, M., Zhao, Y., Qin, B., De Clercq, O., 2016. Semeval-2016 task 5: Aspect based sentiment analysis, ProWorkshop on Semantic Evaluation (SemEval-2016). Association for Computational Linguistics, pp. 19-30.
Poria, S., Cambria, E., Gelbukh, A., 2016. Aspect extraction for opinion mining with a deep convolutional neural network. Knowl.-Based Syst. 108, 42–49.
Ramírez-Tinoco, F.J., Alor-Hernández, G., Sánchez-Cervantes, J.L., Salas-Zárate, M.d.P., Valencia-García, R., 2019. Use of sentiment analysis techniques in healthcare domain. Current Trends in Semantic Web Technologies: Theory and Practice, 189-212.
Remus, R., Quasthoff, U., Heyer, G., 2010. SentiWS-A Publicly Available German-language Resource for Sentiment Analysis, LREC.
Rodríguez-Ibánez, M., Casánez-Ventura, A., Castejón-Mateos, F., Cuenca-Jiménez, P.-M., 2023. A review on sentiment analysis from social media platforms. Expert Systems with Applications, 119862.
Rognone, L., Hyde, S., Zhang, S.S., 2020. News sentiment in the cryptocurrency market: An empirical comparison with Forex. Int. Rev. Financ. Anal. 69, 101462.
Rosenthal, S., Farra, N., Nakov, P., 2019. SemEval-2017 task 4: Sentiment analysis in Twitter. arXiv preprint arXiv:1912.00741.
Rustam, F., Ashraf, I., Mehmood, A., Ullah, S., Choi, G.S., 2019. Tweets classification on the base of sentiments for US airline companies. Entropy 21, 1078.
Saadany, H., Orasan, C., 2020. Is it great or terrible? preserving sentiment in neural machine translation of arabic reviews. arXiv preprint arXiv:2010.13814.
Sankar, H., Subramaniyaswamy, V., 2017. Investigating sentiment analysis using machine learning approach, 2017 International conference on intelligent sustainable systems (ICISS). IEEE 87–92.
Sarkis-Onofre, R., Catalá-López, F., Aromataris, E., Lockwood, C., 2021. How to properly use the PRISMA Statement. Syst. Rev. 10, 1–3.
Shaik, T., Tao, X., Dann, C., Xie, H., Li, Y., Galligan, L., 2023. Sentiment analysis and opinion mining on educational data: A survey. Natural Language Processing Journal 2, 100003.
Shaukat, K., Hameed, I.A., Luo, S., Javed, I., Iqbal, F., Faisal, A., Masood, R., Usman, A., Shaukat, U., Hassan, R., 2020. Domain specific lexicon generation through sentiment analysis. International Journal of Emerging Technologies in Learning (iJET) 15, 190–204.
Shen, Y., Heacock, L., Elias, J., Hentel, K.D., Reig, B., Shih, G., Moy, L., 2023. ChatGPT and other large language models are double-edged swords. Radiological Society of North. America e230163.
Smetanin, S., 2020. The applications of sentiment analysis for Russian language texts: Current challenges and future perspectives. IEEE Access 8, 110693–110719.
Socher, R., Perelygin, A., Wu, J., Chuang, J., Manning, C.D., Ng, A.Y., Potts, C., 2013. Recursive deep models for semantic compositionality over a sentiment treebank. In: Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing, pp. 1631–1642.
Su, J., Chen, Q., Wang, Y., Zhang, L., Pan, W., Li, Z., 2023. Sentence-level sentiment analysis based on supervised gradual machine learning. Sci. Rep. 13, 14500.
Sudirjo, F., Diantoro, K., Al-Gasawneh, J.A., Azzaakiyyah, H.K., Ausat, A.M.A., 2023. Application of ChatGPT in Improving Customer Sentiment Analysis for Businesses. Jurnal Teknologi Dan Sistem Informasi Bisnis 5, 283–288.
Susnjak, T., 2024. Applying bert and chatgpt for sentiment analysis of lyme disease in scientific literature. Methods and Protocols. Springer, Borrelia burgdorferi, pp. 173–183.

Taboada, M., Brooke, J., Tofiloski, M., Voll, K., Stede, M., 2011. Lexicon-based methods for sentiment analysis. Comput. Linguist. 37, 267–307.
Tai, K.S., Socher, R., Manning, C.D., 2015. Improved semantic representations from tree-structured long short-term memory networks. arXiv preprint arXiv:1503.00075.
Tan, X., Cai, Y., Xu, J., Leung, H.-F., Chen, W., Li, Q., 2020. Improving aspect-based sentiment analysis via aligning aspect embedding. Neurocomputing 383, 336–347.
Thakkar, H., Patel, D., 2015. Approaches for sentiment analysis on twitter: A state-of-art study. arXiv preprint arXiv:1512.01043.
Thorp, H.H., 2023. ChatGPT is fun, but not an author. American Association for the Advancement of Science 313.
Tian, Y., Yang, L., Sun, Y., Liu, D., 2021. Cross-domain end-to-end aspect-based sentiment analysis with domain-dependent embeddings. Complexity 2021, 1–11.
Tran, T., Ba, H., Huynh, V.-N., 2019. Measuring hotel review sentiment: An aspect-based sentiment analysis approach, Integrated Uncertainty in Knowledge Modelling and Decision Making: 7th International Symposium, IUKM 2019, Nara, Japan, March 27–29, 2019, Proceedings 7. Springer 393–405.
Verma, S., Saini, M., Sharan, A., 2017. Deep sequential model for review rating prediction. In: 2017 Tenth International Conference on Contemporary Computing (IC3). IEEE, pp. 1–6.
Wang, C., Peng, T., Zhang, Y., Yue, L., Liu, L., 2022. AOPSS: A Joint Learning Framework for Aspect-Opinion Pair Extraction as Semantic Segmentation, Asia-Pacific Web (APWeb) and Web-Age Information Management (WAIM) Joint International Conference on Web and Big Data. Springer, pp. 101-113.
Wang, Z., Xie, Q., Ding, Z., Feng, Y., Xia, R., 2023. Is ChatGPT a good sentiment analyzer? A preliminary study. arXiv preprint arXiv:2304.04339.
Wang, H., Zheng, L., 2016. Sentiment classification of Chinese online reviews: a comparison of factors influencing performances. Enterprise Information Systems 10, 228–244.
Wen, J., Zhang, G., Zhang, H., Yin, W., Ma, J., 2020. Speculative text mining for document-level sentiment classification. Neurocomputing 412, 52–62.
Willmott, C.J., Matsuura, K., 2005. Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. Climate Res. 30, 79–82.
Wilson, T., Wiebe, J., Hoffmann, P., 2005. Recognizing contextual polarity in phrase-level sentiment analysis. Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing 347–354.
Wu, N., 2012. The maximum entropy method. Springer Science & Business Media.
Wu, Y., Rabe, M.N., Hutchins, D., Szegedy, C., 2022. Memorizing transformers. arXiv preprint arXiv:2203.08913.
Wu, L., Morstatter, F., Liu, H., 2018. SlangSD: building, expanding and using a sentiment dictionary of slang words for short-text sentiment classification. Lang. Resour. Eval. 52, 839–852.
Wu, C., Xiong, Q., Yi, H., Yu, Y., Zhu, Q., Gao, M., Chen, J., 2021. Multiple-element joint detection for aspect-based sentiment analysis. Knowl.-Based Syst. 223, 107073.
Xing, F.Z., Cambria, E., Welsch, R.E., 2018. Natural language based financial forecasting: a survey. Artif. Intell. Rev. 50, 49–73.
Xu, Q.A., Chang, V., Jayne, C., 2022b. A systematic review of social media-based sentiment analysis: Emerging trends and challenges. Decision Analytics Journal 3, 100073.
Xu, J., Chen, D., Qiu, X., Huang, X., 2016. Cached long short-term memory neural networks for document-level sentiment classification. arXiv preprint arXiv:1610.04989.
Xu, M., Wang, D., Feng, S., Yang, Z., Zhang, Y., 2022a. Kc-isa: An implicit sentiment analysis model combining knowledge enhancement and context features. In: Proceedings of the 29th International Conference on Computational Linguistics, pp. 6906–6915.
Xu, Q., Zhu, L., Dai, T., Yan, C., 2020. Aspect-based sentiment classification with multi-attention network. Neurocomputing 388, 135–143.
Yadav, A., Vishwakarma, D.K., 2020. Sentiment analysis using deep learning architectures: a review. Artif. Intell. Rev. 53, 4335–4385.
Yan, H., Dai, J., Qiu, X., Zhang, Z., 2021. A unified generative framework for aspect-based sentiment analysis. arXiv preprint arXiv:2106.04300.
Yin, D., Meng, T., Chang, K.-W., 2020. Sentibert: A transferable transformer-based architecture for compositional sentiment semantics. arXiv preprint arXiv:2005.04114.
You, Q., Luo, J., Jin, H., Yang, J., 2015. Robust image sentiment analysis using progressively trained and domain transferred deep networks. Proceedings of the AAAI Conference on Artificial Intelligence.
You, Q., Luo, J., Jin, H., Yang, J., 2016. Cross-modality consistent regression for joint visual-textual sentiment analysis of social multimedia. In: Proceedings of the Ninth ACM International Conference on Web Search and Data Mining, pp. 13–22.
Yousif, A., Niu, Z., Tarus, J.K., Ahmad, A., 2019. A survey on sentiment analysis of scientific citations. Artif. Intell. Rev. 52, 1805–1838.
Zadeh, A., Zellers, R., Pincus, E., Morency, L.-P., 2016. Mosi: multimodal corpus of sentiment intensity and subjectivity analysis in online opinion videos. arXiv preprint arXiv:1606.06259.
Zaremba, W., Sutskever, I., Vinyals, O., 2014. Recurrent neural network regularization. arXiv preprint arXiv:1409.2329.
Zavattaro, S.M., French, P.E., Mohanty, S.D., 2015. A sentiment analysis of US local government tweets: The connection between tone and citizen involvement. Gov. Inf. Q. 32, 333–341.
Zhang, W., Li, X., Deng, Y., Bing, L., Lam, W., 2021. Towards generative aspect-based sentiment analysis. In: Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (volume 2: Short Papers), pp. 504–510.
Zhang, W., Li, X., Deng, Y., Bing, L., Lam, W., 2022a. A survey on aspect-based sentiment analysis: Tasks, methods, and challenges. IEEE Trans. Knowl. Data Eng.
Zhang, Y., Peng, T., Han, R., Han, J., Yue, L., Liu, L., 2022b. Synchronously tracking entities and relations in a syntax-aware parallel architecture for aspect-opinion pair extraction. Appl. Intell. 52, 15210–15225.
Zhao, Q., Ma, S., Ren, S., 2022. KESA: a knowledge enhanced approach for sentiment analysis. arXiv preprint arXiv:2202.12093.
Zhao, A., Yu, Y., 2021. Knowledge-enabled BERT for aspect-based sentiment analysis. Knowl.-Based Syst. 227, 107220.
Zhou, Y., Srikumar, V., 2021. A closer look at how fine-tuning changes BERT. arXiv preprint arXiv:2106.14282.
Zhu, C., Yi, B., Luo, L., 2024. Base on contextual phrases with cross-correlation attention for aspect-level sentiment analysis. Expert Syst. Appl. 241, 122683.
Zunic, A., Corcoran, P., Spasic, I., 2020. Sentiment analysis in health and well-being: systematic review. JMIR Med. Inform. 8, e16023.