



Systematic reviews in sentiment analysis: a tertiary study

Alexander Ligthart¹ · Cagatay Catal² · Bedir Tekinerdogan¹

Accepted: 8 February 2021
© The Author(s) 2021

Abstract

With advanced digitalisation, we can observe a massive increase of user-generated content on the web that provides opinions of people on different subjects. Sentiment analysis is the computational study of analysing people's feelings and opinions for an entity. The field of sentiment analysis has been the topic of extensive research in the past decades. In this paper, we present the results of a tertiary study, which aims to investigate the current state of the research in this field by synthesizing the results of published secondary studies (i.e., systematic literature review and systematic mapping study) on sentiment analysis. This tertiary study follows the guidelines of systematic literature reviews (SLR) and covers only secondary studies. The outcome of this tertiary study provides a comprehensive overview of the key topics and the different approaches for a variety of tasks in sentiment analysis. Different features, algorithms, and datasets used in sentiment analysis models are mapped. Challenges and open problems are identified that can help to identify points that require research efforts in sentiment analysis. In addition to the tertiary study, we also identified recent 112 deep learning-based sentiment analysis papers and categorized them based on the applied deep learning algorithms. According to this analysis, LSTM and CNN algorithms are the most used deep learning algorithms for sentiment analysis.

Keywords Sentiment analysis · Tertiary study · Systematic literature review · Sentiment classification

✉ Cagatay Catal
ccatal@qu.edu.qa

Alexander Ligthart
alexander.ligthart@wur.nl

Bedir Tekinerdogan
bedir.tekinerdogan@wur.nl

¹ Information Technology Group, Wageningen University & Research, Wageningen, The Netherlands

² Department of Computer Science & Engineering, Qatar University, Doha, Qatar

1 Introduction

Sentiment analysis or *opinion mining* is the computational study of people's opinions, sentiments, emotions, and attitudes towards entities such as products, services, issues, events, topics, and their attributes (Liu 2015). As such, sentiment analysis can allow tracking the mood of the public about a particular entity to create actionable knowledge. Also, this type of knowledge can be used to understand, explain, and predict social phenomena (Pozzi et al. 2017). For the business domain, sentiment analysis plays a vital role in enabling businesses to improve strategy and gain insight into customers' feedback about their products. In today's customer-oriented business culture, understanding the customer is increasingly important (Chagas et al. 2018).

The explosive growth of discussion platforms, product review websites, e-commerce, and social media facilitates a continuous stream of thoughts and opinions. This growth makes it challenging for companies to get a better understanding of customers' aggregate opinions and attitudes towards products. The explosion of internet-generated content coupled with techniques like sentiment analysis provides opportunities for marketers to gain intelligence on consumers' attitudes towards their products (Rambocas and Pacheco 2018). Extracting sentiments from product reviews helps marketers to reach out to customers who need extra care, which will improve customer satisfaction, sales, and ultimately benefits businesses (Vyas and Uma 2019).

Sentiment analysis is a multidisciplinary field, including psychology, sociology, natural language processing, and machine learning. Recently, the exponentially growing amounts of data and computing power enabled more advanced forms of analytics. Machine learning, therefore, became a dominant tool for sentiment analysis. There is an abundance of scientific literature available on sentiment analysis, and there are also several secondary studies conducted on the topic.

A secondary study can be considered as a review of primary studies that empirically analyze one or more research questions (Nurdiani et al. 2016). The use of secondary studies (i.e., systematic reviews) in software engineering was suggested in 2004, and the term "Evidence-based Software Engineering" (EBSE) was coined by Kitchenham et al. (2004). Nowadays, secondary studies are widely used as a well-established tool in software engineering research (Budgen et al. 2018). The following two kinds of secondary studies can be conducted within the scope of EBSE:

- Systematic Literature Review (SLR): An SLR study aims to identify relevant primary studies, extract the required information regarding the research questions (RQs), and synthesize the information to respond to these RQs. It follows a well-defined methodology and assesses the literature in an unbiased and repeatable way (Kitchenham and Charters 2007).
- Systematic Mapping Study (SMS): An SMS study presents an overview of a particular research area by categorizing and mapping the studies based on several dimensions (i.e., facets) (Petersen et al. 2008).

SLR and SMS studies are different than traditional review papers (a.k.a., survey articles) because we systematically search in electronic databases and follow a well-defined protocol to identify the articles. There are also several differences between SLR and SMS studies (Catal and Mishra 2013; Kitchenham et al. 2010b). For instance, while RQs of the SLR studies are very specific, RQs of SMS are general. The search process of the SLR is

driven by research questions, but the search process of the SMS is based on the research topic. For the SLR, all relevant papers must be retrieved, and quality assessments of identified articles must be performed; however, requirements for the SMS are less stringent.

When there is a sufficient number of secondary studies on a research topic, a tertiary study can be performed (Kitchenham et al. 2010a; Nurdiani et al. 2016). A tertiary study synthesizes data from secondary studies and provides a comprehensive review of research in a research area (Rios et al. 2018). They are used to summarize the existing secondary studies and can be considered as a special form of review that uses other secondary studies as primary studies (Raatikainen et al. 2019).

Although sentiment analysis has been the topic of some SLR studies, a tertiary study characterizing these systematic reviews has not been performed yet. As such, the aim of our study is to identify and characterize systematic reviews in sentiment analysis and present a consolidated view of the published literature to better understand the limitations and challenges of sentiment analysis. We follow the research methodology guidelines suggested for the tertiary studies (Kitchenham et al. 2010a).

The objective of this study is thus to better understand the sentiment analysis research area by synthesizing results of these secondary studies, namely SLR and SMS, and providing a thorough overview of the topic. The methodology that we followed applies a systematic literature review to a sample of systematic reviews, and therefore, this type of tertiary study is valuable to determine the potential research areas for further research.

As part of this tertiary study, different models, tasks, features, datasets, and approaches in sentiment analysis have been mapped and also, challenges and open problems in this field are identified. Although tertiary studies have been performed for other topics in several fields such as software engineering and software testing (Raatikainen et al. 2019; Nurdiani et al. 2016; Verner et al. 2014; Cruzes and Dybå, 2011; Cadavid et al. 2020), this is the first study that performs a tertiary study on sentiment analysis.

The main contributions of this article are three-fold:

- We present the results of the first tertiary study in the literature on sentiment analysis.
- We identify systematic review studies of sentiment analysis systematically and explain the consolidated view of these systematic studies.
- We support our study with recent survey papers that review deep learning-based sentiment analysis papers and explain the popular lexicons in this field.

The rest of the paper is organized as follows: Sect. 2 provides the background and related work. Section 3 explains the methodology, which was followed in this study. Section 4 presents the results in detail. Section 5 provides the discussion, and Sect. 6 explains the conclusions.

2 Background and related work

Sentiment analysis and opinion mining are often used interchangeably. Some researchers indicate a subtle difference between sentiments and opinions, namely that opinions are more concrete thoughts, whereas sentiments are feelings (Pozzi et al. 2017). However, sentiment and opinion are related constructs, and both sentiment and opinion are included when referring to either one. This research adopts *sentiment analysis* as a general term for both opinion mining and sentiment analysis.

Sentiment analysis is a broad concept that consists of many different tasks, approaches, and types of analysis, which are explained in this section. In addition, an overview of sentiment analysis is represented in Fig. 1, which is adapted from (Hemmatian and Sohrabi 2017; Kumar and Jaiswal 2020; Mite-Baidal et al. 2018; Pozzi et al. 2017; Ravi and Ravi 2015). Cambria et al. (2017) stated that a holistic approach to sentiment analysis is required, and only categorization or classification is not sufficient. They presented the problem as a three-layer structure that includes 15 Natural Language Processing (NLP) problems as follows:

- Syntactics layer: Microtext normalization, sentence boundary disambiguation, POS tagging, text chunking, and lemmatization
- Semantics layer: Word sense disambiguation, concept extraction, named entity recognition, anaphora resolution, and subjectivity detection

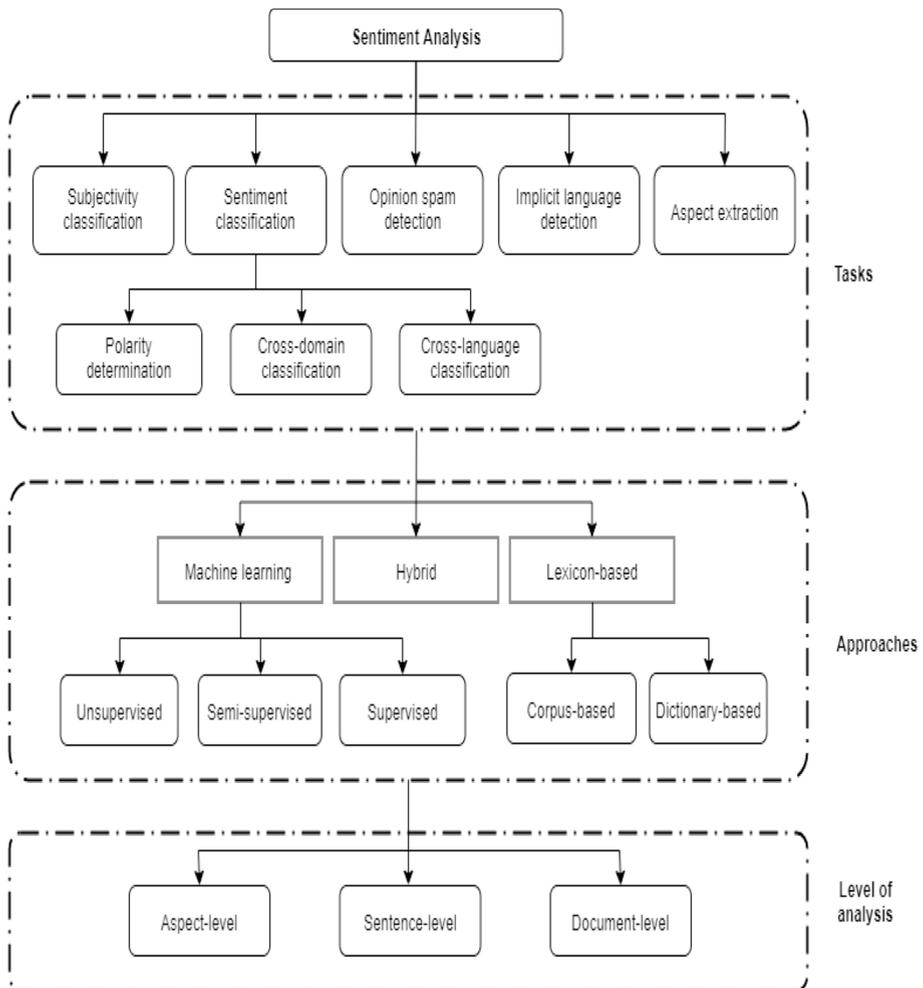


Fig. 1 Sentiment analysis concept overview

- Pragmatics layer: Personality recognition, sarcasm detection, metaphor understanding, aspect extraction, and polarity detection

Cambria (2016) state that approaches for sentiment analysis and affective computing can be divided into the following three categories: knowledge-based techniques, statistical approaches (e.g., machine learning and deep learning approaches), and hybrid techniques that combine the knowledge-based and statistical techniques.

Sentiment analysis models can adopt different pre-processing methods and apply a variety of feature selection methods. While pre-processing means transforming the text into normalized tokens (e.g., removing article words and applying the stemming or lemmatization techniques), feature selection means determining what features will be used as inputs. In the following subsections, related tasks, approaches, and levels of analysis are presented in detail.

2.1 Tasks

2.1.1 Sentiment classification

One of the most widely known and researched tasks in sentiment analysis is sentiment classification. Polarity determination is a subtask of sentiment classification and is often improperly used when referring to sentiment analysis. However, it is merely a subtask aimed at identifying sentiment polarity in each text document. Traditionally, polarity is classified as either positive or negative (Wang et al. 2014). Some studies include a third class called *neutral*. Cross-domain and cross-language classification are subtasks of sentiment classification that aim to transfer knowledge from a data-rich source domain to a target domain where data and labels are limited. The cross-domain analysis predicts the sentiment of a target domain, with a model (partly) trained on a more data-rich source domain. A popular method is to extract domain invariant features whose distribution in the source domain is close to that of the target domain (Peng et al. 2018). The model can be extended with target domain-specific information. The cross-language analysis is practiced in a similar way by training a model on a source language dataset and testing it on a different language where data is limited, for example by translating the target language to the source language before processing (Can et al. 2018). Xia et al. (2015) stated that opinion-level context is beneficial to solve polarity ambiguity of sentiment words and applied the Bayesian model. Word polarity ambiguity is one of the challenges that need to be addressed for sentiment analysis. Vechtomova (2017) showed that the information retrieval-based model is an alternative to machine learning-based approaches for word polarity disambiguation.

2.1.2 Subjectivity classification

Subjectivity classification is a task to determine the existence of subjectivity in the text (Kasmuri and Basiron 2017). The goal of subjectivity classification is to restrict unwanted objective data objects for further processing (Kamal 2013). It is often considered the first step in sentiment analysis. Subjectivity classification detects *subjective clues*, words that carry emotion or subjective notions like ‘expensive’, ‘easy’, and ‘better’ (Kasmuri and Basiron 2017). These clues are used to classify text objects as subjective or objective.

2.1.3 Opinion spam detection

The growing popularity of e-commerce websites and review websites caused opinion spam detection to be a prominent issue in sentiment analysis. Opinion spams also referred to as false or fake reviews are intelligently written comments that either promote or discredit a product. Opinion spam detection aims to identify three types of features that relate to a fake review: review content, metadata of review, and real-life knowledge about the product (Ravi and Ravi 2015). Review content is often analyzed with machine learning techniques to uncover deception. Metadata includes the star rating, IP address, geo-location, user-id, etc.; however, in many cases, it is not accessible for analysis. The third method includes real-life knowledge. For instance, if a product has a good reputation, and suddenly the inferior product is rated superior in some period, reviews of that period might be suspected.

2.1.4 Implicit language detection

Implicit language refers to humor, sarcasm, and irony. There are vagueness and ambiguity in this form of speech, which is sometimes hard to detect even for humans. However, an implicit meaning to a sentence can completely flip the polarity of a sentence. Implicit language detection often aims at understanding facts related to an event. For example, in the phrase “I love pain”, pain is a factual word with a negative polarity load. The contradiction of the factual word ‘pain’ and subjective word ‘love’ can indicate sarcasm, irony, and humor. More traditional methods for implicit language detection include exploring clues such as emoticons, expressions for laughter, and heavy punctuation mark usage (Filatova 2012).

2.1.5 Aspect extraction

Aspect extraction refers to retrieving the target entity and aspects of the target entity in the document. The target entity can be a product, person, event, organization, etc. (Akshi Kumar and Sebastian 2012). People’s opinions on various parts of a product need to be identified for fine-grained sentiment analysis (Ravi and Ravi 2015). Aspect extraction is especially important in sentiment analysis of social media and blogs that often do not have predefined topics.

Multiple methods exist for aspect extraction. The first and most traditional method is frequency-based analysis. This method finds frequently used nouns or compound nouns (POS tags), which are likely to be aspects. A rule of thumb that is often used is that if the (compound) noun occurs in at least 1% of the sentences, it is considered an aspect. This straightforward method turns out to be quite powerful (Schouten and Frasincar 2016). However, there are some drawbacks to this method (e.g., not all nouns are referring to aspects).

Syntax-based methods find aspects by means of syntactic relations they are in. A simple example is identifying aspects that are preceded by a modifying adjective that is a sentiment word. This method allows for low-frequency aspects to be identified. The drawback of this method is that many relations need to be found for complete coverage, which requires knowledge of sentiment words. Extra aspects can be found if more sentiment words that serve as adjectives can be identified. Qiu et al. (2009) propose a syntax-based algorithm

that identifies aspects as well as sentiment words that works both ways. The algorithm identifies sentiment words for known aspects and aspects for known sentiment words.

2.2 Approaches

2.2.1 Machine learning-based approaches

Machine learning approaches for sentiment analysis tasks can be divided into three categories: unsupervised learning, semi-supervised learning, and supervised learning.

The unsupervised learning methods group unlabelled data into clusters that are similar to each other. For example, the algorithm can consider data as similar based on common words or word pairs in the document (Li and Liu 2014).

Semi-supervised learning uses both labeled and unlabelled data in the training process (da Silva et al. 2016a, b). A set of unlabelled data is complemented with some examples of labeled data (often limited) included building a classifier. This technique can yield decent accuracy and requires less human effort compared to supervised learning. In cross-domain and cross-language classification, domain, or language invariant features can be extracted with the help of unlabelled data, while fine-tuning the classifier with labeled target data (Peng et al. 2018). Semi-supervised learning is especially popular for Twitter sentiment analysis, where large sets of unlabelled data are available (da Silva et al. 2016a, b). Hus-sain and Cambria (2018) compared the computational complexity of several semi-supervised learning methods and presented a new semi-supervised model based on biased SVM (bSVM) and biased Regularized Least Squares (bRLS). Wu et al. (2019) developed a semi-supervised Dimensional Sentiment Analysis (DSA) model using the variational autoencoder algorithm. DSA calculates the sentiment score of texts based on several dimensions, such as dominance, valence, and arousal. Xu and Tan (2019) proposed the target-oriented semi-supervised sequential generative model (TSSGM) for target-oriented aspect-based sentiment analysis and showed that this approach outperforms two semi-supervised learning methods. Han et al. (2019) developed a semi-supervised model using dynamic thresholding and multiple classifiers for sentiment analysis. They evaluated their model on the Large Movie Review dataset and showed that it provides higher performance than the other models. Duan et al. (2020) proposed the Generative Emotion Model with Categorized Words (GEM-CW) model for stock message sentiment classification and demonstrated that this model is effective. Gupta et al. (2018) investigated the semi-supervised approaches for low resource sentiment classification and showed that their proposed methods improve the model performance against supervised learning models.

The most widely known machine learning method is supervised learning. This approach trains a model with labeled source data. The trained model can subsequently make predictions for an output considering new unlabelled input data. In most cases, supervised learning often outperforms unsupervised and semi-supervised learning approaches, but the dependency on labeled training data can require lots of human effort and is therefore sometimes inefficient (Hemmatian and Sohrabi 2017).

Machine learning methods are increasingly popular for aspect extraction. The most commonly used approach for aspect extraction is *topic modeling*, an unsupervised method that assumes any document contains a certain amount of hidden topics (Hemmatian and Sohrabi 2017). Latent Dirichlet Allocation (LDA) algorithm, which has many different variations, is a popular topic modeling algorithm (Nguyen and Shirai 2015) that allows observations to be explained by unsupervised grouping of similar data. LDA outputs some

topics of a text document and attributes each word in the document to one of the identified topics. The drawback of machine learning methods is that they require lots of labeled data.

2.2.2 Deep learning-based approaches

Deep learning is a sub-branch of machine learning that uses deep neural networks. Recently, deep learning algorithms have been widely applied for sentiment analysis. In this section, first, we discuss the articles that present an overview of papers that applied deep learning for sentiment analysis. These articles are neither SLR nor SMS papers. Instead, they are either traditional review (a.k.a., survey) articles or comparative assessment papers that explain the existing deep learning-based approaches in addition to the experimental analysis. Later, we also present some of the deep learning-based models used in sentiment analysis papers.

In Table 1, we present the survey papers that analyzed deep learning-based sentiment analysis papers. In this table, we also show the number of papers investigated in these survey papers.

Dang et al. (2020) presented a summary of 32 deep learning-based sentiment analysis papers and analyzed the performance of Deep Neural Networks (DNN), Convolutional Neural Networks (CNN), and Recurrent Neural Networks (RNN) on eight datasets. They selected these deep learning algorithms because they are the most widely used deep learning algorithms according to their analysis of 32 deep learning-based sentiment analysis papers. They used both word embedding and term frequency-inverse document frequency (TF-IDF) to prepare inputs for classification algorithms and reported that the RNN-based model using word embedding achieved the best performance among other algorithms. However, the processing time of the RNN-based model is ten times larger than the CNN-based one. In addition, they reported that the following deep learning algorithms were used in the 32 deep learning-based sentiment analysis papers: CNN, Long-Short Term Memory (LSTM) (tree-LSTM, discourse-LSTM, coattention-LSTM, bi-LSTM), Gated Recurrent Units (GRU), RNN, Coattention-MemNet, Latent Rating Neural Network (LRNN), Simple Recurrent Networks (SRN), and Recurrent Neural Tensor Network (RNTN)).

Yadav and Vishwakarma (2019) reviewed 130 research papers that apply deep learning techniques in sentiment analysis. They identified the following deep learning methods used for sentiment analysis: CNN, Recursive Neural Network (Rec NN), RNN (LSTM and GRU), Deep Belief Networks (DBN), Attention-based Network, Bi-RNN, and Capsule Network. They reported that LSTM provides better results, and the use of deep learning approaches for sentiment analysis is promising. However, they stated that they require a huge amount of data, and there is a lack of training datasets.

Zhang et al. (2018) published a survey article on the application of deep learning methods for sentiment analysis. They explained several papers that address one of the following levels: document level, sentence level, and the aspect level sentiment classification. The applied algorithms per analysis level are listed as follows:

- Document-level sentiment classification: Artificial Neural Networks (ANN), Stacked Denoising Autoencoder (DSA), Denoising Autoencoder, CNN, LSTM, GRU, Memory Network, and GRU-based Encoder
- Sentence-level sentiment classification: CNN, RNN, Semi-supervised Recursive Autoencoders Network (RAE), Recursive Neural Network, Recursive Neural Tensor

Table 1 Survey articles that investigated the use of deep learning in sentiment analysis

Retrieved From	References	Title	# of papers	Year
MDPI	Dang et al. (2020)	Sentiment analysis based on deep learning: A comparative study	32	2020
Springer	Yadav and Vishwakarma (2019)	Sentiment analysis using deep learning architectures: a review	130	2019
Wiley	Zhang et al. (2018)	Deep learning for sentiment analysis: A survey	62*	2018
Wiley	Rojas-Barahona (2016)	Deep learning for sentiment analysis. Language and Linguistics Compass	8	2016
Springer	Habimana et al. (2020a)	Sentiment analysis using deep learning approaches: an overview	54*	2020
Science Direct	Do et al. (2019)	Deep learning for aspect-based sentiment analysis: a comparative review	40	2019
Arxiv.org	Minaee et al. (2020)	Deep learning based text classification: A comprehensive review	22	2020

*Document-level sentiment analysis, sentence-level sentiment analysis, aspect-level sentiment analysis, multi-modal data for sentiment analysis, multi-lingual sentiment analysis papers are counted

Network, Dynamic CNN, LSTM, CNN-LSTM, Bi-LSTM, and Recurrent Random Walk Network

- Aspect-level sentiment classification: Adaptive Recursive Neural Network, LSTM, Bi-LSTM, Attention-based LSTM, Memory Network, Interactive Attention Network, Recurrent Attention Network, and Dyadic Memory Network

Rojas-Barahona (2016) presented an overview of deep learning approaches used for sentiment analysis and divided the techniques into the following categories:

- Non-Recursive Neural Networks: RNN (variant: Bi-RNN), LSTM (variant: Bi-LSTM), and CNN (variants: CNN-Multichannel, CNN-non-static, Dynamic CNN)
- Recursive Neural Networks: Recursive Autoencoders and Constituency Tree Recursive Neural Networks
- Combination of Non-Recursive and Recursive Methods: Tree-Long Short-Term Memory (Tree-LSTM) and Deep Recursive Neural Networks (Deep RsNN)

For the movie reviews dataset, Rojas-Barahona (2016) showed that the Dynamic CNN model provides the best performance. For the Sentiment TreeBank dataset, the Constituency Tree-LSTM that is a Recursive Neural Network outperforms all the other algorithms.

Habimana et al. (2020a) reviewed papers that applied deep learning algorithms for sentiment analysis and also performed several experiments with the specified algorithms on different datasets. They reported that dynamic sentiment analysis, sentiment analysis for heterogeneous information, and language structure are the main challenges for the sentiment analysis research field. They categorized the techniques used in the papers based on several analysis levels that are listed as follows:

- Document-level Sentiment Analysis: CNN-based models, RNN with attention-based models, RNN with the user and product attention-based models, Adversarial Network Models, and Hybrid Models
- Sentence-Level Sentiment Classification: Unsupervised Pre-Trained Networks (UPN), CNN, Recurrent Neural Networks, Deep Reinforcement Learning (DRL), RNN, RNN with cognition attention-based models
- Aspect-based Sentiment Analysis: Attention-based models with aspect information, attention-based models with the aspect context, RNN with attention memory model, RNN with commonsense knowledge model, CNN-based model, and Hybrid model

Do et al. (2019) presented an overview of over 40 deep learning approaches used for aspect-based sentiment analysis. They categorized papers based on the following categories: CNN, RNN, Recursive Neural Network, and Hybrid methods. Also, they presented the advantages, disadvantages, and implications for aspect-based sentiment analysis (ABSA). They concluded that deep learning and ABSA are still in the early stages, and there are four main challenges in this field, namely domain adaptation, multi-lingual application, technical requirements (labeled data and computational resources and time), and linguistic complications.

Minaee et al. (2020) reviewed more than 150 deep learning-based text classification studies and presented their strengths and contributions. 22 of these studies proposed approaches for sentiment analysis. They provided more than 40 popular text classification datasets and showed the performance of some deep learning models on popular datasets. Since they did not only focus on sentiment analysis problems, they explained

other kinds of models used for other tasks such as news categorization, topic analysis, question answering (QA), and natural language inference. They explained the following deep learning models in their paper: Feed-forward neural networks, RNN-based models, CNN-based models, Capsule Neural Networks, Models with attention mechanism, Memory augmented networks, Transformers, Graph Neural Networks, Siamese Neural Networks, Hybrid models, Autoencoders, Adversarial training, and Reinforcement learning. The challenges reported in this study are new datasets for multi-lingual text classification, interpretable deep learning models, and memory-efficient models. They concluded that the use of deep learning in text classification improves the performance of the models.

Some of the highly cited deep learning-based sentiment analysis papers are shown in Table 2.

Kim (2014) performed several experiments with the CNN algorithm for sentence classification and showed that even with little parameter tuning, the CNN model that includes only one convolutional layer provides better performance than the state-of-the-art models of sentiment analysis.

Wang et al. (2016) developed an attention-based LSTM approach that can learn aspect embeddings. These aspects are used to compute the attention weights. Their models provided a state-of-the-art performance on SemEval 2014 dataset. Similarly, Pergola et al. (2019) proposed a topic-dependent attention model for sentiment classification and showed that the use of recurrent unit and multi-task learning provides better representations for accurate sentiment analysis.

Chen et al. (2017) developed the Recurrent Attention on Memory (RAM) model and showed that their model outperforms other state-of-the-art techniques on four datasets, namely SemEval 2014 (two datasets), Twitter dataset, and Chinese news comment dataset. Multiple attentions were combined with a Recurrent Neural Network in this study.

Ma et al. (2018) incorporated a hierarchical attention mechanism to the LSTM network and also extended the LSTM cell to incorporate commonsense knowledge. They demonstrated that the combination of this new LSTM model called Sentic LSTM and the attention architecture outperforms the other models for targeted aspect-based sentiment analysis.

Chen et al. (2016) developed a hierarchical LSTM model that incorporates user and product information via different levels of attention. They showed that their model achieves significant improvements over models without user and product information on IMDB, Yelp2013, and Yelp2014 datasets.

Wehrmann et al. (2017) proposed a language-agnostic sentiment analysis model based on the CNN algorithm, and the model does not require any translation. They demonstrated that their model outperforms other models on a dataset, including tweets from four languages, namely English, German, Spanish, and Portuguese. The dataset consists of 1.6 million annotated tweets (i.e., positive, negative, and neutral) from 13 European languages.

Ebrahimi et al. (2017) presented the challenges of building a sentiment analysis platform and focused on the 2016 US presidential election. They reported that they reached the best accuracy using the CNN algorithm, and the content-related challenges were hashtags, links, and sarcasm.

Poria et al. (2018) investigated three deep learning-based architectures for multimodal sentiment analysis and created a baseline based on state-of-the-art models.

Xu et al. (2019) developed an improved word representation approach, used the weighted word vectors as input into the Bi-LSTM model, obtained the comment text representation, and applied the feedforward neural network classifier to predict the comment sentiment tendency.

Table 2 Highly cited deep learning-based sentiment analysis papers

Retrieved From	Reference	Title	# of citations	Year
Arxiv.org	Kim (2014)	Convolutional neural networks for sentence classification	9645	2014
Google Scholar	Wang et al. (2016)	Attention-based LSTM for aspect-level sentiment classification	919	2016
Google Scholar	Chen et al. (2017)	Recurrent attention network on Memory for aspect sentiment analysis	347	2017
Google Scholar	Ma et al. (2018)	Targeted aspect-based sentiment analysis via embedding commonsense knowledge into an attentive LSTM	268	2018
Google Scholar	Chen et al. (2016)	Neural sentiment classification with user and product attention	227	2016
IEEE	Wehrmann et al. (2017)	A character-based convolutional neural network for language-agnostic Twitter sentiment analysis	83	2017
IEEE	Ebrahimi et al. (2017)	Challenges of sentiment analysis for dynamic events	60	2017
IEEE	Poria et al. (2018)	Multimodal sentiment analysis: Addressing key issues and setting up the baselines	53	2018
IEEE	Xu et al. (2019)	Sentiment analysis of comment texts based on BiLSTM	46	2019
IEEE	Majumder et al. (2019)	Sentiment and sarcasm classification with multitask learning	46	2019

Majumder et al. (2019) proposed a GRU-based Neural Network that can be trained on sarcasm or sentiment datasets. They demonstrated that multitask learning-based approaches provide better performance than standalone classifiers developed on sarcasm and sentiment datasets.

After investigating these above-mentioned survey and highly cited articles, we searched in Google Scholar by using our search criteria (i.e., “deep learning” and “sentiment analysis”) to reach the recent state-of-the-art deep learning-based studies published in 2020. We retrieved 112 deep learning-based sentiment analysis papers published in 2020 and extracted the applied deep learning algorithms from these papers. In Appendix (Table 16), we present these recent deep learning-based sentiment analysis papers. In Table 3, we show the distribution of applied deep learning algorithms used in these 112 recent papers.

According to this table, the most applied algorithm is the LSTM algorithm (i.e., 35.53%) and the second most used algorithm is CNN (i.e., 33.33%). The other widely used algorithms are GRU (i.e., 8.77%) and RNN (i.e., 7.89%) algorithms. However, the other well-known deep learning algorithms such as DNN, Recursive Neural Network (ReNN), Capsule Network (CapsN), Generative Adversarial Network (GAN), Deep Q-Network, and Autoencoder have not been preferred much and used only in a few studies. Most of the hybrid approaches also combined the CNN and LSTM algorithms and therefore, they were represented under these categories. As this analysis indicates, most of the recent deep learning-based studies followed the supervised learning machine learning approach.

Table 3 Distribution of deep learning papers published in 2020

ID	Algorithm	# of usages	Percentage (%)
1	LSTM	81	35.53
2	CNN	76	33.33
3	GRU	20	8.77
4	RNN	18	7.89
5	Bidirectional Encoder Representations from Transformers (BERT)	7	3.07
6	DNN	4	1.75
7	ReNN	4	1.75
8	Graph Convolutional Neural Network (GCN)	3	1.32
9	Capsule Network (CapsN)	2	0.88
10	Recurrent Convolutional Neural Network (RCNN)	2	0.88
11	Distillation Network (DN)	2	0.88
12	Generative Adversarial Network (GAN)	1	0.44
13	Gated Alternate Neural Network (GANN)	1	0.44
14	Category Attention Network (CAN)	1	0.44
15	Recurrent Memory Neural Network (ReMemNN)	1	0.44
16	Interactive Rule Attention Network (IRAN)	1	0.44
17	Self-Attention based Hierarchical Dilated Convolutional Neural Network (SA-HDCNN)	1	0.44
18	Fusion-Extraction Network (FENet)	1	0.44
19	Deep Q-Network	1	0.44
20	Autoencoder	1	0.44
	Total	228	100

2.2.3 Lexicon-based approaches

The traditional approach for sentiment analysis is the lexicon-based approach (Hemmatian and Sohrabi 2017). Lexicon-based methods scan through the documents for words that express positive or negative feelings to humans. Negative words would be ‘bad’, ‘ugly’, ‘scary’ while positive words are, for example, ‘good’ or ‘beautiful’. The values of these words are documented in a lexicon. Words with high positive or negative values are mostly adjectives and adverbs. Sentiment analysis shows to be extremely dependent on the domain of interest (Vinodhini 2012). For example, analyzing movie reviews can yield very different results compared to analyzing Twitter data due to different forms of language used. Therefore, the lexicon used for sentiment analysis needs to be adjusted according to the domain of interest. This can be a time-consuming process. However, lexicon-based methods do not require training data, which is a big advantage (Shayaa et al. 2018).

There are two main approaches to creating sentiment lexicons: dictionary-based and corpus-based. The dictionary-based approach starts with a small set of sentiment words, and iteratively expands the lexicon with synonyms and antonyms from existing dictionaries. In most cases, the dictionary-based approach works best for general purposes. Corpus-based lexicons can be tailored to specific domains. The approach starts with a list of general-purpose sentiment words and discovers other sentiment words from a domain corpus based on co-occurring word patterns (Mite-Baidal et al. 2018).

2.2.4 Hybrid approaches

There are different hybrid approaches in the literature. Some of them aim to extend machine learning models with lexicon-based knowledge (Behera et al. 2016). The goal is to combine both methods to yield optimal results using an effective feature set of both lexicon and machine learning-based techniques (Munir Ahmad et al. 2017). This way, the deficiencies and limitations of both approaches can be overcome.

Recently, researchers focused on the integration of symbolic and subsymbolic Artificial Intelligence (AI) for sentiment analysis (Cambria et al. 2020). Machine learning (also, deep learning) is considered to be a bottom-up approach and applies subsymbolic AI. This is extremely useful for exploring a huge amount of data and discovering interesting patterns in the data. Although this type of bottom-up approach works quite well for image classification tasks, they are not very effective for natural language processing tasks. For effective communication, we learn many issues such as cultural awareness and commonsense in a top-down manner instead of a bottom-up manner (Cambria et al. 2020). Therefore, these researchers applied subsymbolic AI (i.e., deep learning) to recognize patterns in text and represented them in a knowledge base using symbolic AI (i.e., logic and semantic networks). They built a new commonsense knowledge base called SenticNet for the sentiment analysis problem and concluded that coupling symbolic AI and subsymbolic AI is crucial to passing to the natural language understanding stage from natural language processing.

Minaee et al. (2019) developed an ensemble model using LSTM and CNN algorithm and demonstrated that this ensemble model provides better performance than the individual models.

2.2.5 Milestones of sentiment analysis research

Recently, Poria et al. (2020) investigated the challenges and new research directions in sentiment analysis research. Also, they presented the key milestones of sentiment analysis for the last two decades. We adapted their timeline figure for the last decade. In Fig. 2, we present the most promising works of sentiment analysis research. For a more detailed illustration of milestones, we refer the readers to the article of Poria et al. (2020).

2.3 Levels of analysis

Sentiment analysis can be implemented at the following three levels: document, sentence, and aspect level. We elaborate on these in the next paragraphs.

2.3.1 Document-level

Document-level analysis considers the whole text document as a unit of analysis (Wang et al. 2014). It is a simplified task that presumes that the entire document originates from a single opinion holder. Document analysis comes with some issues, namely that there could be multiple and mixed opinions in a document expressed in many different ways, sometimes with implicit language (Akshi Kumar and Sebastian 2012). Typically, documents are revised on a sentence or aspect level before determining the polarity of the entire text document.

2.3.2 Sentence-level:

Sentence-level analysis considers specific sentences in a text and is especially used for subjectivity classification. Text documents typically consist of sentences that either contain opinion or not. Subjectivity classification analyses individual sentences in a document to detect whether the sentence contains facts or emotions and opinions. The main goal of

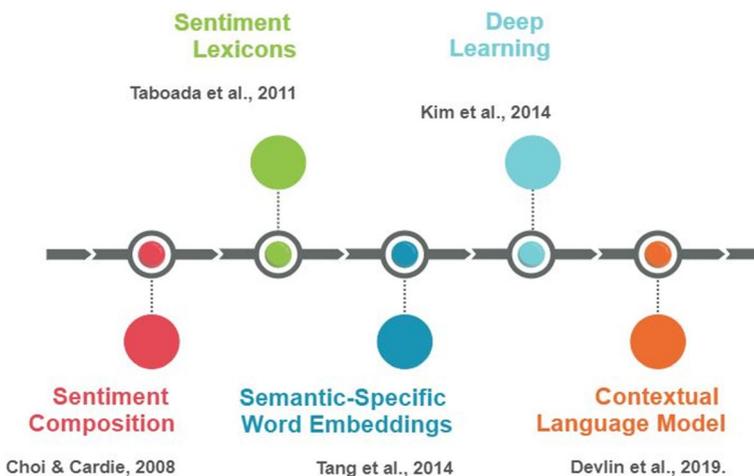


Fig. 2 Milestones of sentiment analysis research for the last decade

subjectivity classification is to exclude sentences that do not contain sentiment or opinion (Akshi Kumar and Sebastian 2012). This analysis often includes subjectivity classification as a step to either include or exclude sentences for analysis.

2.3.3 Aspect-level

Aspect-level analysis is a challenging topic in sentiment analysis. It refers to analyzing sentiments about specific entities and their aspects in a text document, not merely the overall sentiment of the document (Tun Thura Thet et al. 2010). It is also known as entity-level or feature-level analysis. Even though the general sentiment of a document may be classified as positive or negative, the opinion holder can have a divergent opinion about specific aspects of an entity (Akshi Kumar and Sebastian 2012). In order to measure aspect-level opinion, aspects of the entity need to be identified. Valdivia et al. (2017) stated that aspect-based sentiment analysis is beneficial to the business manager because customer opinions are extracted in a transparent way. Also, they reported that ironic expression detection in TripAdvisor is still an open problem and also, labeling of reviews should not only focus on user ratings because some users write positive sentences on negative user ratings and vice versa. Poria et al. (2016) proposed a new algorithm called Sentic LDA (Latent Dirichlet Allocation) and improved the LDA algorithm with semantic similarity for aspect-based sentiment analysis. They concluded that this new algorithm helps researchers to pass to the semantics analysis from the syntactical analysis in aspect-based sentiment analysis by using the common-sense computing (Cambria et al. 2009) and improves the clustering process (Poria et al. 2016).

2.4 Popular lexicons

Several survey articles discussed the popular lexicons used in sentiment analysis. Dang et al. (2020) reported the following popular sentiment analysis lexicons in their article: Sentiment 140, Tweets Airline, Tweets Semeval, IMDB Movie Reviews (1), IMDB Movie Reviews (2), Cornell Movie Reviews, Book Reviews, and Music Reviews datasets. Habimana et al. (2020a) explained the following popular lexicons in their survey article: IMDB, IMDB2, SST-5, SST-2, Amazon, SemEval 2014-D1, SemEval 2014-D2, SemEval 2017, STS, STS-Gold, Yelp, HR (Chinese), MR, Sanders, Deutsche Bahn (Deutsch), ASTD (Arabic), YouTube, CMU-MOSI, and CMU-MOSEI. Do et al. (2019) reported the following datasets widely used in sentiment analysis papers: Customer review data, SemEval 2014, SemEval 2015, SemEval 2016, ICWSM 2010 JDPA Sentiment Corpus, Darmstadt Service Review Corpus, FiQA ABSA, and target-dependent Twitter sentiment classification dataset. Minaee et al. (2020) explained the following datasets used for sentiment analysis: Yelp, IMDB, Movie Review, SST, MPQA, Amazon, and aspect-based sentiment analysis datasets (SemEval 2014 Task-4, Twitter, and SentiHood). Researchers who would like to perform a new study are suggested to look at these articles because links and other details per lexicon are presented in detail in these articles.

2.5 Advantages, disadvantages, and performance of the models

Several studies have been performed to compare the performance of existing models for sentiment analysis. Each model has its own advantages and weaknesses. For the aspect-based sentiment analysis, Do et al. (2019) divided models based on the following three

categories: CNN, RNN, and Recurrent Neural Networks. The advantages of CNN-based models are fast computation, the ability to extract local patterns and represent non-linear dynamics. The disadvantage of the CNN-based model is the high demand for data. The advantages of RNN-based models are that they do not require a huge amount of data, they have a distributed hidden state that stores previous computations, and they require fewer parameters. The disadvantages are that they cannot capture long-term dependencies, and they select the last hidden state to represent the sentence. The advantages of Recurrent Neural Networks are their simple architectures and their ability to learn tree structures. The disadvantages are that they require parsers that might be slow, and they are still at early stages. It was reported that RNN-based models provide better performance than CNN-based models, and more research is required for Recurrent Neural Networks.

Yadav and Vishwakarma (2019) reported that deep learning-based models are gaining popularity for different sentiment analysis tasks. They stated that CNN followed by LSTM (an RNN algorithm) provides the highest accuracy for document-level sentiment classification, researchers focused on RNN algorithms (particularly, LSTM) for sentence-level sentiment classification and aspect-level sentiment classification, and RNN models the best-performing ones for multi-domain sentiment classification. They also discussed the merits and demerits of CNN, Recursive Neural Networks (RecNN), RNN, LSTM, GRU, DBN models.

The advantage of DBN is the ability to learn the dimension of vocabulary using different layers. The disadvantages of DBN are that they are computationally expensive and unable to remember the previous task.

The advantage of GRU is that it is computationally less expensive, it has a less complex structure, and it can capture interdependencies between sentences. The disadvantage of GRU is that it does not have a memory unit, and its performance is lower than the LSTM model on larger datasets.

The advantage of LSTM is that they perform better than CNN, they can extract sequential information, and they can forget/remember things selectively. The disadvantage of LSTM is that it is considerably slower, each output should be reconciled to a sentence, and it is computationally expensive.

The advantage of RNN models is that they provide better performance than CNN models, have fewer parameters, and capture long-distance dependency features. The disadvantage of RNN models is that they cannot process long sequences.

The advantage of CNN models is that they are less expensive in terms of computational complexity and faster compared to RNN, LSTM, and GRU algorithms. Also, they can discover relevant features from different parts of the word. The disadvantage of LSTM models is that they cannot preserve long-term dependency and ignores this type of long-distance features.

The advantage of RecNN is that they are good at learning hierarchical structure and therefore, they provide better performance for NLP tasks. The disadvantage of RecNN models is that their efficiency is dramatically affected in the case of informal data that do not have grammatical rules and training can be difficult because structure changes for every sample.

Despite the excellent performance of deep learning models, there are some drawbacks. The following drawbacks are discussed by Yadav and Vishwakarma (2019):

- A huge amount of data is required to train the models and finding these large datasets is not easy in many cases

- They work like a black box, it is hard to understand how they predict the sentiment of the text
- The performance of the models is affected by the hyperparameters and the selection of these hyperparameters is very challenging
- Training time is very long and most of the time they require GPU support and large RAM

Yadav and Vishwakarma (2019) performed experiments to compare the execution time and accuracy of several deep learning algorithms. They reported that the LSTM algorithm and its variations such as Bi-LSTM and GRU require long training and execution time compared to other deep learning models. However, these LSTM-based algorithms better performance. Therefore, there is a trade-off between time and accuracy parameters when selecting the deep learning model.

3 Methodology

In this section, the methodology of our tertiary study is presented. This study can be considered as a systematic review study that targets secondary studies on sentiment analysis, which is a widely researched topic. There are several reviews and mapping studies available on sentiment analysis in the literature. In this section, we focus on synthesizing the results of these secondary studies. Hence, we conduct a tertiary study. The study design is based on the systematic literature review (SLR) protocol suggested by Kitchenham and Charters (2007) and the format followed by the tertiary study papers of Curcio et al. (2019); Raatikainen et al. (2019). This study reviews two types of secondary studies:

- SLR: These studies are performed to aggregate results related to specific research questions.
- SMS: These studies aim to find and classify primary studies in a specific research topic. This method is more explorative compared to the SLR and is used to identify available literature prior to undertaking an SLR.

Both are considered secondary studies as they review primary studies. A pragmatic comparison between SLR and SMS is discussed by Kitchenham et al. (2011). Three main phases for conducting this research are *planning, conducting, and reporting* the review (Kitchenham 2004). Planning refers to identifying the need for the review and developing the review protocol. The goal of this tertiary study is to gather a broad overview of the current state of the art in sentiment analysis and to identify open problems and challenges in the field.

3.1 Research questions

The following research questions have been defined for this study:

- *RQ1* What are the adopted features (input/output) in sentiment analysis?
- *RQ2* What are the adopted approaches in sentiment analysis?
- *RQ3* What domains have been addressed in the adopted data sets?
- *RQ4* What are the challenges and open problems with respect to sentiment analysis?

3.2 Search process

This section provides insight into the process of determining secondary studies to include. Not all databases are equally relevant to this research topic. Databases that are used to identify secondary studies are adopted from the search strategy of secondary studies on sentiment analysis (Genc-Nayebi and Abran 2017; Hemmatian and Sohrabi 2017; Kumar and Jaiswal 2020; Sharma and Dutta 2018). The following databases are included in this study: *IEEE*, *Science Direct*, *ACM*, *Springer*, *Wiley*, and *Scopus*. To find the relevant literature, databases are searched for the title, abstract, and keywords based on the following query:

(“sentiment analysis” OR “sentiment classification” OR “opinion mining”) AND (“SLR” OR “systematic literature review” OR “systematic mapping” OR “mapping study”)

This query results in 43 hits. As stated before, this study only considers systematic literature reviews and systematic mapping studies since they are considered of higher quality and more in-depth compared to survey articles. Inclusion and exclusion criteria are formulated, as shown in Table 4.

All secondary studies are analyzed and classified according to the inclusion and exclusion criteria in Table 4. After this process, 16 secondary studies are selected.

3.3 Quality assessment

The confidence placed in the secondary studies is on the quality assessment of the articles. For a tertiary study, the quality assessment is especially important (Goulão et al. 2016). The DARE criteria proposed by York University Centre for Reviews and Dissemination (CDR) and adopted in this study are often used in the context of software engineering (Goulão et al. 2016; Rios et al. 2018; Curcio et al. 2019; Goulão et al. 2016; Kitchenham et al. 2010a). The criteria are based on four questions (CQs), as shown in Table 5. For each selected article, the criteria are scored based on a three-point scale, as described in Table 6, adopted from (Kitchenham et al. 2010a, b).

The scoring procedure is Yes = 1, Partial = 0.5, and No = 0. The assessment is conducted by the researchers. The results of the quality assessment are shown in Table 7. Two studies are excluded based on the results, leaving a total amount of 14 studies remaining for analysis.

Table 4 Inclusion and exclusion criteria

<i>Inclusion criteria</i>	
IC1:	Systematic literature review studies
IC2:	Systematic mapping studies
IC3:	Studies that include machine learning models
IC4:	Studies published from 2009 until the present
<i>Exclusion criteria</i>	
EC1:	Secondary studies not directly related to sentiment analysis of textual data
EC2:	Primary studies on sentiment analysis
EC3:	Traditional review articles (‘surveys’)
EC4:	Studies not considered as secondary studies

Table 5 Quality criteria questions

Quality criteria questions

- CQ1: Are reviews' inclusion and exclusion criteria described and appropriate?
 CQ2: Is the literature search likely to have covered all the relevant studies?
 CQ3: Did the user reviewers assess the quality/validity of the included studies?
 CQ4: Were the basic data/studies adequately described?

Table 6 Quality criteria scoring metrics

Scoring metrics

- CQ1: Y (yes), the inclusion criteria are explicitly defined in the paper, P (Partly), the inclusion criteria are implicit; N (no), the inclusion criteria are not defined and cannot be readily inferred
 CQ2: Y, the authors have either searched four or more digital libraries and included additional search strategies or identified and referenced all journals addressing the topic of interest; P, the authors have searched 3 or 4 digital libraries with no extra search strategies, or searched a defined but restricted set of journals and conference proceedings; N, the authors have searched up to 2 digital libraries or an extremely restricted set of journals. Note that scoring question 2 also requires the evaluator to consider whether the digital libraries were appropriate for the specific SLR
 CQ3: Y, the authors have explicitly defined quality criteria and extracted them from each primary study; P, the research question involves quality issues that are addressed by the study; N no explicit quality assessment of individual papers has been attempted, or quality data has been extracted but not used
 CQ4: Y, Information is presented about each paper so that the data summaries can clearly be traced to relevant papers; P, only summary information is presented about individual papers, e.g., papers are grouped into categories, but it is not possible to link individual studies to each category; N, the results of the individual studies are not specified, i.e., the individual primary studies are not cited

3.4 Additional data

In order to provide an overview of the selected secondary studies, Table 8 shows the following data extracted from the articles: Research focus, number of primary studies included in the review, year of publication, paper type (conference/journal/book chapter), and source. In addition, an overview of the research questions of the secondary studies is provided, as shown in Table 9. The reference numbers in Table 8 are used throughout the rest of this paper.

4 Results

This section addresses the results of the research questions derived from 14 secondary studies. For each research question, tables with aggregate results and in-depth descriptions and interpretations are presented. The selected secondary studies discuss specific sentiment analysis tasks. It is important to note that different tasks in sentiment analysis require different features and approaches. Therefore, a brief overview of each paper is presented. Note that in-depth analysis and synthesis of the articles are presented later in this section.

Table 7 Quality assessment results

	Author	Title	Type	Q1	Q2	Q3	Q4	Total
1	Genç-Nayebi and Abbran (2017)	A systematic literature review: Opinion mining studies from mobile app store user reviews	SLR	Y	Y	Y	Y	4
2	Al-Moslmi et al. (2017)	Approaches to Cross-Domain Sentiment Analysis: A Systematic Literature Review	SLR	Y	Y	N	Y	3
3	Qazi et al. (2017)	A systematic literature review on opinion types and sentiment analysis techniques: Tasks and challenges	SLR	Y	Y	Y	Y	4
4	Ibrahim and Salim (2013)	Opinion analysis for twitter and Arabic tweets: a systematic literature review	SLR	Y	Y	Y	Y	4
5	Shayaa et al. (2018)	Sentiment analysis of big data: Methods, applications, and open challenges	SLR	Y	P	N	Y	2.5
6	Kumar and Sharma (2017)	Systematic literature review on opinion mining of big data for government intelligence	SLR	Y	Y	N	Y	3
7	Ahmad et al. (2018)	Sentiment analysis using SVM: A systematic literature review	SLR	Y	N	Y	P	2.5
8	Kumar and Jaiswal (2020)	Systematic literature review of sentiment analysis on Twitter using soft computing techniques	SLR	Y	Y	P	Y	3.5
9	Kumar and Garg (2019)	Systematic literature review on context-based sentiment analysis in social multimedia	SLR	Y	Y	Y	Y	4
10	Kasmuri and Basiron (2017)	Subjectivity analysis in opinion mining—A systematic literature review	SLR	Y	Y	Y	P	3.5
11	Madhala et al. (2018)	Systematic literature review on customer emotions in social media	SLR	Y	Y	N	Y	3
12	Mite-Baidal et al. (2018)	Sentiment analysis in education domain: A systematic literature review	SLR	P	Y	N	Y	2.5
13	Ragini and Anand (2016)	Sentiment analysis: A comprehensive overview and the state of art research challenges	SLR	N	N	N	P	0.5
14	Salah et al. (2019)	A systematic review on opinion mining and sentiment analysis in social media	SLR	P	Y	N	Y	2.5
15	Sharma and Dutta (2018)	Polarity Determination of Movie Reviews: A Systematic Literature Review	SLR	P	N	P	Y	1.5
16	de Oliveira et al. (2018)	Mining on Line General Opinions About Sustainability of Hotels: A Systematic Literature Mapping	SMS	Y	Y	N	Y	3

Table 8 Data extraction of selected articles

Ref. no.	Author	Research focus	# studies included	Year	Paper type	Source
1	Genc-Nayebi and Abran (2017)	Identify approaches	24	2017	Journal	Journal of systems and software
2	Al-Moslmi et al. (2017)	Identify cross-domain approaches	28	2017	Journal	IEEE access
3	Qazi et al. (2017)	Identify approaches and challenges	24	2017	Journal	Internet research
4	Ibrahim and Salim (2013)	Identify approaches for twitter opinion mining	65	2013	Journal	Journal of theoretical and applied information technology
5	Shayaa et al. (2018)	Identify approaches, application domains and challenges	58	2018	Journal	IEEE access
6	Kumar and Sharma (2017)	Identify approaches for government intelligence	194	2017	Journal	Webology
7	Ahmad et al. (2018)	Identify SVM applications and trends	8	2018	Journal	International journal of advanced computer science and application
8	Kumar and Jaiswal (2020)	Identify approaches for twitter opinion mining	60	2019	Journal	Concurrency and computation: practice and experience
9	Kumar and Garg (2019)	Identify context-based approaches and challenges	37	2019	Journal	Multimedia tools and applications
10	Kasmuri and Basiron (2017)	Identify approaches and performance of subjectivity analysis	97	2017	Journal	International journal of advances in soft computing and its applications
11	Madhala et al. (2018)	Identify emotion classification approaches and challenges	23	2018	Conference paper	5th European conference on social media
12	Mite-Baidal et al. (2018)	Identify approaches in education domain	8	2018	Conference paper	International conference on technologies and innovation
13	Salah et al. (2019)	Identify content-based approaches and challenges	20	2019	Journal	International journal of information systems
14	de Oliveira et al. (2018)	Identify approaches on hotel review analysis	29	2018	Conference paper	International conference on computational science and its applications

Table 9 Research questions per article

Ref. no.	Title	Research questions
1	A systematic literature review: Opinion mining studies from mobile app store user reviews	<ul style="list-style-type: none"> (1) Which specific data mining techniques are used for reviews on software distribution platforms? (2) How do the studies remedy the ‘domain dependency’ challenge for app store reviews? (3) What criteria make a review useful? (4) How can spam reviews be differentiated from legitimate reviews? (5) Does the study extract targeted/desired software features from application reviews?
2	Approaches to Cross-Domain Sentiment Analysis: A Systematic Literature Review	<ul style="list-style-type: none"> (1) For a given source-target domain, are the cross-domain approaches able to perform as well as they do in in-domain sentiment analysis? (2) Are the cross-domain algorithms sensitive to variations in data representation? (3) Does the cross-domain result depend on whether the source and target domains are heterogeneous or homogeneous?
3	A systematic literature review on opinion types and sentiment analysis techniques: Tasks and challenges	<ul style="list-style-type: none"> (1) How many opinion types are present in online reviews according to published studies? (2) Which are the tasks resolved by machine learning techniques for SA of opinions? (3) Which are the Sentic computing techniques that resolved the challenges for SA of opinions?

Table 9 (continued)

Ref. no.	Title	Research questions
4	Opinion analysis for twitter and Arabic tweets: a systematic literature review	<p>What are the techniques and features that can be used for Arabic opinion mining in a tweet?</p> <p>(1) What are the features and techniques that have been used for the subjectivity identification of tweets? Are these features and techniques used for Arabic tweets?</p> <p>(2) What are the features and techniques that have been used for polarity identification of tweets? Are these features and techniques used for Arabic tweets?</p> <p>(3) What are the features and techniques that have been used for opinion target identification of tweets? Are these features and techniques used for Arabic tweets?</p> <p>(4) What are the fuzzy aspects that have been studied for twitter opinion mining? Are they used for Arabic tweets?</p> <p>(5) What are the Arabic corpora that have been used in Arabic opinion mining? Are they used for twitter?</p> <p>(6) What are the metrics that have been used to measure the effectiveness of twitter opinion mining in terms of subjectivity, polarity, and opinion target?</p>
5	Sentiment analysis of big data: Methods, applications, and open challenges	<p>(1) What are the trends of opinion mining and sentiment analysis publications from 2000–2016?</p> <p>(2) Which are the areas in which opinion mining and sentiment analysis have been applied?</p> <p>(3) What are the sources of data in the areas in which opinion mining and sentiment analysis have been applied?</p>

Table 9 (continued)

Ref. no.	Title	Research questions
6	Systematic literature review on opinion mining of big data for government intelligence	<ul style="list-style-type: none"> (1) What are the different application areas and techniques of opinion mining? (2) How much work has been done in the area of government intelligence using opinion mining? (3) Which (big) data sets are used for opinion mining in government intelligence? (4) Which techniques of opinion mining have been used in advocacy so far?
7	Sentiment analysis using SVM: A systematic literature review	<ul style="list-style-type: none"> (1) Which are the latest research trends in the domain of sentiment analysis? (2) Which machine learning/lexicon/hybrid technique is considered for comparison with SVM? (3) Which areas of sentiment analysis are considered for investigation by the researchers? (4) Which factors affect the classification results? (5) Which type of dataset is used for performance evaluation?
8	Systematic literature review of sentiment analysis on Twitter using soft computing techniques	<ul style="list-style-type: none"> (1) Which are the most distinguished and relevant journals? (2) On which datasets and domains the studies using SC techniques for SA on Twitter have been conducted? (3) Which are the most frequently used SC techniques for achieving efficient results for the SA on Twitter? (4) What are the widely used performance metrics to evaluate the applied techniques? (5) What is the trend and impact of using SC techniques for SA on Twitter in the past decade?
9	Systematic literature review on context-based sentiment analysis in social multimedia	<ul style="list-style-type: none"> (1) What is the need for using 'context' in sentiment analysis? (2) What are the ways in which 'context' has been used in sentiment analysis? (3) Which are the techniques used for context-based sentiment analysis to date? (4) Which are the social media in which context-based sentiment analysis has been applied to date?

Table 9 (continued)

Ref. no.	Title	Research questions
10	Subjectivity analysis in opinion mining—A systematic literature review	<ol style="list-style-type: none"> (1) What are the common tasks in subjectivity analysis? (2) What are the techniques used to identify subjectivity? (3) What is the corpus used as data sets in subjectivity analysis? (4) What is the technique to represent the subjectivity clues in the analyzed text? (5) What is the performance of the technique that successfully identifies subjectivity? (6) What are the strengths and weaknesses of the technique? (7) What are the affecting elements of the performance of the technique? (8) What are the missing elements in subjectivity analysis?
11	Systematic literature review on customer emotions in social media	<ol style="list-style-type: none"> (1) How is social media data on consumer emotions currently collected? (2) How are large datasets used in detecting emotions? (3) What types of computational methods are being used to detect consumer emotions? (4) What are the limitations and the accuracy of current computational methods in detecting consumer emotions?
12	Sentiment analysis in education domain: A systematic literature review	<ol style="list-style-type: none"> (1) What is the sentiment analysis process? (2) What approaches and digital educational resources are used in sentiment analysis? (3) What are the main benefits of using sentiment analysis in the education domain?
13	A systematic review on opinion mining and sentiment analysis in social media	<ol style="list-style-type: none"> (1) What are the current issues in user content based mining? (2) What are the most appropriate mechanisms used in OM and sentiments analysis? And what are the widest mechanisms used? (3) What are the best feature extraction and selection approaches for OM and SA?

Table 9 (continued)

Ref. no.	Title	Research questions
14	Mining on Line General Opinions About Sustainability of Hotels: A Systematic Literature Mapping	<ol style="list-style-type: none"> <li data-bbox="212 165 256 852">(1) What are the most used text mining techniques to detect the opinion of consumers in online reviews? <li data-bbox="279 165 323 852">(2) Are the online text mining techniques used for detecting the consumers' satisfaction regarding the sustainable practices by the hotels? <li data-bbox="346 165 364 852">(3) What are the countries that have more researchers publishing on the theme? <li data-bbox="388 165 405 852">(4) What years had more publications in the area? <li data-bbox="429 165 446 852">(5) What are the main periodicals and conferences on the theme?

1. Genc-Nayebi and Abran (2017) identify mobile app store opinion mining techniques. Their paper is mainly focused on statistical data mining techniques based on manual classification and correlation analysis. Some machine learning algorithms are discussed in the context of cross-domain analysis and app aspect-extraction. Some interesting challenges in sentiment analysis are proposed.
2. Al-Moslmi et al. (2017) review the cross-domain sentiment analysis. Specific algorithms for cross-domain sentiment analysis are described.
3. Qazi et al. (2017) research the opinion types and sentiment analysis. Opinion types are classified into the following three categories: regular, comparative, and suggestive. Several supervised machine learning techniques are used. Sentiment classification algorithms are mapped.
4. Ahmed Ibrahim and Salim (2013) perform sentiment analysis of Arabic tweets. Their study is focused on mapping features and techniques used for general sentiment analysis.
5. Shayaa et al. (2018) research the big data approach to sentiment analysis. A solid overview of machine learning methods and challenges is presented.
6. A. Kumar and Sharma (2017) research sentiment analysis for government intelligence. Techniques and datasets are mapped.
7. M. Ahmad et al. (2018) focus their research on SVM classification. SVM is the most used machine learning technique in sentiment classification.
8. A. Kumar and Jaiswal (2020) discuss soft computing techniques for sentiment analysis on Twitter. Soft computing techniques include machine learning techniques. Deep learning (CNN in particular) is mentioned as upcoming in recent articles. KPIs are described thoroughly.
9. A. Kumar and Garg (2019) research context-based sentiment analysis. They stress the importance of subjectivity in sentiment analysis and show that deep learning offers opportunities for context-based sentiment analysis.
10. Kasmuri and Basiron (2017) research the subjectivity analysis. Its purpose is to determine whether the text is subjective or objective with objective clues. Subjectivity analysis is a classification problem, and thus, machine learning algorithms are widely used.
11. Madhala et al. (2018) research customer emotion analysis. They review articles that classify emotions from 4 to 51 different classes.
12. Mite-Baidal et al. (2018) research sentiment analysis in the education domain. E-learning is upcoming, and due to the online nature, lots of review data is generated on forums of MOOCS and social media.
13. Salah et al. (2019) research the social media sentiment analysis. Mainly twitter data is used because of the high dimensionality (e. g., retweets, location, user followers no.) and structure.
14. De Oliveira Lima et al. (2018) research opinion mining of hotel reviews specifically aimed at sustainability practices aspects. Limited information on used features is available. The following sections dive into the different models that are used in sentiment analysis, including adopted features, approaches, and datasets.

4.1 RQ1 “What are the adopted features in sentiment analysis?”

Table 10 depicts the common input and output features that articles present for the sentiment analysis approach. Checkmarks indicate that the features are explicitly discussed in the referred article. Traditional approaches commonly use Bag-Of-Words (BOW) method. BOW counts the words, referred to as *n-grams*, in the text and creates a sparse vector with 1 s for present words and 0 s for absent words. These vectors are used as input to machine learning models. N-grams are sets of words that occur next to each other that are combined into one feature. This way, the order of words will be maintained when the text is vectorized. Part-Of-Speech (POS) tags provide feature tags for similar words with a different part of speech in the context. The term frequency-inverse document frequency (TF-IDF) method highlights words or word pairs that often occur in one document but are low in frequency in the entire text corpus. Negation is an important feature to include in lexicon-based approaches. Negation means contradicting or denying something, which can flip the polarity of an opinion or sentiment.

Word embeddings are often used as feature learning techniques in deep learning models. Word embeddings are dense vectors with real numbers for a word or sentence in the text based on the context of the word in the text corpus. This approach, although considered promising, is only discussed to a limited extent in the selected articles.

Output variables differ per sentiment analysis task. Output classes of the identified secondary studies are polarity, subjectivity, emotions classes, or spam identification. Polarity indicates the extent to which the input is considered positive or negative in sentiment. In most cases, the output is classified in a binary way, either positive or negative. Some models include a neutral class as well. Multiple classes of polarity are shown to drastically reduce performance (Al-Moslmi et al. 2017) and are, therefore, not frequently used. One study (Madhala et al. 2018) focuses specifically on emotion classification, with up to 51 different classes of emotions. Some studies (Ahmed Ibrahim and Salim 2013; Kas-muri and Basiron 2017) include subjectivity analysis as part of sentiment analysis. Finally, spam detection is an important task in sentiment analysis, referring to extracting illegitimate means for a review. Examples of spam are untruthful opinions, reviews on the brand instead of on the product, and non-reviews like advertisements and random questions or text (Jindal and Liu 2008).

A clear pattern exists in the use of input and output features. Traditional machine learning models commonly use unigrams and n-grams as input. Variable features are TF-IDF values and POS-tags. Not every feature extraction method is as effective in differing domains. Combinations of input features are often made to reach better performance. Word embeddings are upcoming input features. The most recent articles (Kumar and Garg 2019; Kumar and Jaiswal 2020) explicitly discuss them. Text classification with word embeddings as input is considered a promising technique that is often combined with deep learning methods like recurrent neural networks. The output shows a similar pattern with common and variable features. The common feature is polarity, and variable output features include emotions, subjectivity, and spam type.

Table 10 Overview of common features

Ref. no.	Input						Output					
	Word embed- dings	N-Grams	TF-IDF	POS tags	Negation	Combined features	Polarity	Subjectivity	Emotions	Spam		
1	✓	✓	✓	✓			✓			✓		
2	✓	✓	✓	✓		✓	✓			✓		
3	✓	✓				✓	✓					
4	✓	✓	✓	✓		✓	✓	✓				
5	✓	✓			✓		✓					
6	✓	✓			✓		✓			✓		
7	✓	✓	✓	✓	✓	✓	✓	✓		✓		
8	✓	✓				✓	✓					
9	✓	✓			✓	✓	✓					
10	✓	✓			✓	✓	✓	✓				
11	✓	✓	✓	✓		✓	✓		✓			
12	✓	✓				✓	✓					
13	✓	✓	✓	✓	✓	✓	✓			✓		
14	✓	✓				✓	✓			✓		

Table 11 Common approaches per task in sentiment analysis

	Subjectivity classification	Sentiment classification	Opinion spam detection	Object and aspect extraction	Implicit language detection
Deep Learning		✓			✓
Traditional ML	✓	✓	✓	✓	✓
Unsupervised				✓	
Lexicon	✓	✓	✓	✓	✓
Hybrid		✓			

4.2 RQ2 “What are the adopted approaches in sentiment analysis?”

Different tasks in sentiment analysis require different approaches. Therefore, it is important to note which task requires which approach. Table 11 shows the categories that are used throughout different sentiment analysis tasks.

Table 12 depicts the commonly used approaches for sentiment analysis per selected paper. Machine learning algorithms, including deep learning (DL), unsupervised learning, and ensemble learning, are widely used for sentiment analysis tasks, as well as lexicon-based and hybrid methods. Checkmarks indicate that approaches are explicitly discussed in the referred article. Results are divided into five categories with specific subcategories. Each category with corresponding subcategories are described as follows:

4.2.1 Deep learning

Deep learning models are complex architectures with multiple layers of neural networks to progressively extract high-level features from input. CNN uses convolutional filters to

Table 12 Overview of common approaches in sentiment analysis

Ref. no.	Deep learning		Traditional ML							Lexicon	Hybrid		
	CNN	RNN	SVM	NB	NN	LogR	ME	KNN	RF			DT	
1			✓	✓		✓					✓	✓	
2			✓	✓	✓							✓	
3			✓	✓			✓	✓				✓	
4			✓	✓			✓	✓				✓	✓
5			✓	✓						✓		✓	✓
6			✓	✓	✓		✓			✓		✓	✓
7			✓	✓		✓			✓			✓	✓
8	✓		✓	✓	✓	✓		✓		✓		✓	
9	✓	✓	✓	✓	✓	✓	✓		✓	✓		✓	✓
10			✓	✓		✓				✓		✓	
11		✓	✓	✓								✓	✓
12		✓	✓	✓			✓					✓	✓
13			✓	✓	✓		✓			✓		✓	✓
14			✓	✓		✓				✓		✓	

recognize patterns in data. CNN is widely used in image recognition and, to a lesser extent, in the field of NLP. RNN is designed for recognizing sequential patterns. RNN is especially powerful in cases where context is critical. For this reason, RNN is very promising in sentiment analysis. LSTM networks are a special kind of RNN, that is capable of learning long-term context and dependencies. LSTM is especially powerful in NLP, where long-term dependencies are often important. The discussed deep learning algorithms are considered promising techniques and able to boost the performance of NLP tasks (Socher et al. 2013).

4.2.2 Traditional machine learning

Traditional ML algorithms are still widely used in all kinds of sentiment analysis tasks, including sentiment classification. While deep learning is a promising field, in many cases, traditional ML performs sufficiently well or even better for a specific task compared to deep learning methods, usually on smaller datasets. The traditional supervised machine learning algorithms are Support Vector Machines (SVM), Naive Bayes (NB), Neural Networks (NN), Logistic Regression (LogR), Maximum Entropy (ME), k-Nearest Neighbor (kNN), Random Forest (RF), and Decision Trees (DT).

4.2.3 Lexicon-based

Lexicon-based learning is a traditional approach to sentiment analysis. Lexicon-based methods scan through the documents for words that express positive or negative feelings to humans. Words are defined in a lexicon beforehand, so no learning data is required for this approach.

4.2.4 Hybrid models

In the context of sentiment classification, hybrid models combine the lexicon-based approach with machine learning techniques (Behera et al. 2016) to create a lexicon-enhanced classifier. Lexicons are used for defining domain-related features that are used as input for a machine learning classifier.

4.2.5 Ensemble classification

Ensemble classifiers approach adopts multiple learning algorithms to obtain better performance (Behera et al. 2016). Three main types of ensemble classification methods are bagging (bootstrap aggregating), boosting, and stacking. The bagging method independently learns homogeneous algorithms with data points randomly picked from the training set, following a deterministic averaging process. Boosting learns homogeneous algorithms in a sequential and adaptive way before following an averaging process. Stacking learns heterogeneous classifiers in parallel and combines them to predict an output. An overview of ensemble classifiers is shown in Table 13.

Support Vector Machines (SVM) is the dominant algorithm in the field of sentiment classification. All selected papers include SVM for classification purposes, and in most cases, this technique yields the best performance. Naive Bayes is the second most used algorithm and is praised for its high performance despite the simplicity of the technique. Besides these two dominant algorithms, methods like NN, LogR, ME, kNN, RF, and DT

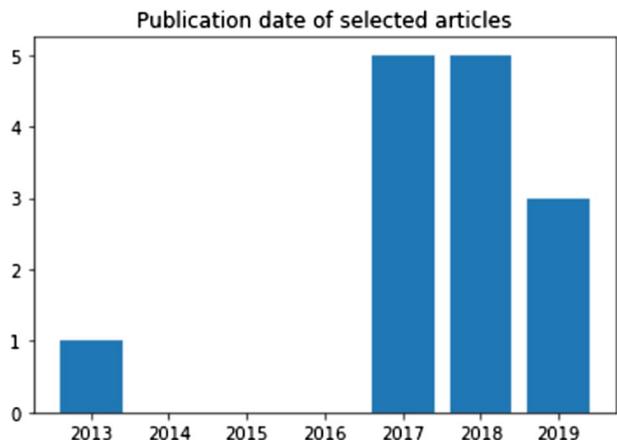
Table 13 Overview of ensemble classification methods

Ref. no.	Boosting	Bagging	Stacking
1			
2			✓
3			
4		✓	✓
5	✓		
6			
7	✓	✓	
8	✓	✓	✓
9			
10			
11			
12	✓		✓
13			
14	✓		

are used throughout different tasks of sentiment analysis. A popular unsupervised approach for aspect extraction is LDA. Hybrid approaches to sentiment classification have been effective by using domain-specific knowledge to create extra features that enhance the performance of the model. Ensemble and hybrid methods often improve the performance and reliability of predictions.

Deep learning algorithms are rising techniques in sentiment analysis. Especially, RNNs and the more complex RNN architecture, LSTM, are increasing in popularity. Even though deep learning is promising for increasing the performance of NLP and sentiment analysis models (Al-Moslmi et al. 2017; Kumar and Garg 2019; Kumar and Jaiswal 2020; Socher et al. 2013), the selected papers only discuss deep learning to a limited extent. The papers that discuss deep learning algorithms are recent papers published in 2018 and 2019, which stresses that sentiment analysis is a timely research subject and that the state-of-the-art is evolving rapidly. Figure 3 shows the year-wise distribution of selected articles. Except for one study from 2013, all selected studies are published in 2017, 2018, and 2019.

Fig. 3 Publication dates of the selected articles



4.3 RQ3 “What domains have been addressed in the adopted data sets?”

Datasets for sentiment analysis are typically user-generated textual content. The text differs a lot depending on the domain and platform that the content is derived from. For example, social media data is usually very subjective and full of informal speech, whereas news article websites are mostly objective and formally written. Twitter data is limited to a certain number of characters and contains hashtags and references, whereas product review websites take a specific product into account and describe this in-depth. ML models trained on a specific domain provide poor performance when tested on a dataset from a different domain. Different domains have different language use and, therefore, require different methods for analysis. Table 14 depicts the domains of the adopted datasets per study. Checkmarks indicate that datasets from the domain are explicitly mentioned in the referred article.

Social media data is the most widely used source of data. This data is usually easy to obtain through APIs. Especially, tweets are popular because they are relatively similar in format (e.g., a limited number of characters). Twitter has an API where tweets can be scraped on specific subjects, time range, hashtags, etc. Tweets contain worldwide real-time information on entities. Furthermore, scraped tweets contain information about the location, number of retweets, number of likes, and much more. Some reviewed articles focus specifically on Twitter data (Ahmed Ibrahim and Salim 2013; Kumar and Jaiswal 2020). Other social media platforms like Facebook and Tumblr are also used for sentiment analysis.

Reviews of products, hotels, and movies are also commonly used for text classification models. Reviews are usually combined with a star rating (i.e., label), which makes them suitable for machine learning models. Star ratings indicate polarity. This way, no labor-intensive manual labeling process or predefined lexicon is required.

Table 14 Overview of domains of the adopted datasets

Ref. no.	Movies	E-commerce	Traveling	News and media	Social media	Mobile apps	Multi-domain
1						✓	
2		✓	✓	✓	✓		✓
3	✓	✓			✓		
4					✓		
5	✓	✓	✓	✓	✓		
6	✓	✓		✓	✓		
7		✓	✓		✓		
8					✓		
9	✓		✓		✓		
10		✓	✓	✓	✓		
11			✓		✓		
12					✓		
13					✓		
14		✓	✓		✓		

4.4 RQ4 “What are the challenges and open problems with respect to sentiment analysis?”

All of the 14 selected papers include challenges and open problems in sentiment analysis. Table 15 shows the challenges that are explicitly described in the papers. These challenges are categorized and sorted by the number of selected papers that explicitly mention the challenge.

Domain dependency is a well-known challenge in sentiment analysis; most of the models that we build are dependent on the domain it was built in. Linguistic dependency is the second most stated and well-known challenge that originates from the same deeper problem. Specific text corpora per domain or language need to be available for the optimal performance of the ML model. Some studies investigate multi-lingual or multi-domain models.

Most papers use English text corpora. Spanish and Chinese are the second most used languages in sentiment analysis. Limited literature is available in other languages. Some studies attempted to create a multi-language model (Al-Moslmi et al. 2017), but this is still a challenging task (Kumar and Garg 2019; Qazi et al. 2017). Multi-lingual systems are an interesting topic for further research.

Deep learning is a promising but complex technique where syntactic structures and word order can be retained. Deep learning still poses some challenges and is not widely researched in the selected articles. Opinion spam or fake review detection is a prominent issue in sentiment analysis where the internet has become an integral part of life, and false information spreads just as fast as accurate information on the web (Vosoughi et al. 2018). Another major challenge is the multi-class classification. In general, more output classes in a classifier reduce the performance (Al-Moslmi et al. 2017). Multiple polarity classes and multiple classes of emotions (Madhala et al. 2018) have been shown to dramatically reduce the performance of the model.

Further challenges are incomplete information, implicit language, typos, slang, and all other kinds of inconsistencies in language use. Combining text with corresponding pictures, audio, and video is also challenging.

5 Discussion

The goal of this study is to present an overview of the current application of machine learning models and corresponding challenges in sentiment analysis. This is done by critically analyzing the selected secondary studies and extracting the relevant data considering the predefined research questions. This tertiary study follows the guidelines proposed by Kitchenham and Charters (2007) for conducting systematic literature reviews. The study initially selected 16 secondary studies. After the quality assessment, 14 secondary papers remained for data extraction. The research methodology is transparent and designed in such a way that it can be reproduced by other researchers. Like any secondary study, there are also some limitations to this tertiary study.

The SLRs included in this study have their specific research focus on sentiment analysis. Even though the methodology of the 14 secondary studies is similar, the documentation of techniques and methods differs a lot. Besides that, some SLR papers are more comprehensive than others. This made the data extraction process harder and prone to mistakes.

Table 15 challenges in sentiment analysis

Challenge	Description	# Refs	References
1 Domain dependency	Differing context, domain jargon, cultural factors and language use between domains hinder performance and require training data for the domain of interest	10	Genc-Nayebi and Abran (2017), Al-Moslmi et al. (2017), Qazi et al. (2017), Ibrahim and Salim (2013), Shayaa et al. (2018), Kumar and Sharma (2017), Kumar and Garg (2019), Madhala et al. (2018), Mite-Baidal et al. (2018), Salah et al. (2019)
2 Language dependency	Different languages require different training data. Multilingual models are challenging	7	Genc-Nayebi and Abran (2017), Al-Moslmi et al. (2017), Qazi et al. (2017), Ibrahim and Salim (2013), Kumar and Garg (2019), Kasnuri and Basiron (2017), Salah et al. (2019)
3 Deep Learning	Deep learning with word embeddings offers opportunities but is least explored in literature	3	Al-Moslmi et al. (2017), Kumar and Jaiswal (2020), Kumar and Garg (2019)
4 Opinion Spam Detection	Opinion spam or fake review detection is one of the largest problems in sentiment analysis. Current systems are limited and not yet mature	2	Genc-Nayebi and Abran (2017), Kumar and Sharma (2017)
5 Sparsity and incomplete information	As data can be incomplete, models should be able to detect and supplement missing information for more accurate results	2	Shayaa et al. (2018), Kumar and Sharma (2017)
6 Emotion detection	Detecting different emotions is challenging due to the many different emotion classes and implicit communication	2	Qazi et al. (2017), Kumar and Jaiswal (2020)
7 Multi-source data fusion	Analysis of an event from different sources and then constructing semantic associations between text, image, and video data is challenging but possibly rewarding	2	Shayaa et al. (2018), Kumar and Jaiswal (2020)
8 Typo's, slang, and informal speech	Typo's, slang and informal speech cause increased dimensionality and are harder to decode. Occurs often on forums and social media	2	Kumar and Sharma (2017), Kumar and Jaiswal (2020)
9 Implicit language	Implicit language like humor, sarcasm, and irony is hard to detect	2	Kumar and Jaiswal (2020), Kumar and Garg (2019)
10 Multi-class classification	Complementing the output with three or more sentiment classes results in a severe decrease in sentiment analysis performance and is challenging	1	Al-Moslmi et al. (2017)

Another limitation concerns the selection process. The criteria for inclusion are restricted to SLR and SMS papers. Some other studies chose to include non-systematic literature reviews as well to complement results, but we did not include traditional survey papers because they do not systematically synthesize the papers in a field.

The first threat to validity is related to the inclusion criteria for methods in research questions. Checkmarks in the tables of RQ2, RQ3, and RQ4 are placed when something is explicitly mentioned in the referred paper. The included secondary studies have their specific research focus with different sentiment analysis tasks and corresponding machine learning approaches. For instance, Kasmuri and Basiron (2017) discuss subjectivity classification, which typically uses different approaches compared to other sentiment analysis tasks. This variation in research focus influences the checkmarks placed in the tables.

Another threat related to inclusion criteria is that some secondary studies have more included papers than others. For example, Kumar and Sharma (2017) included 194 primary studies, where Mite-Baidal et al. (2018) only included eight primary studies. It is likely that papers with a higher number of included primary articles mention more different techniques and challenges, and thus, more checkmarks are placed in the tables compared to papers with a lower number of primary articles included.

Lastly, this tertiary study only considers the selected secondary papers and does not consult the primary papers selected by the secondary papers. If any mistakes are made in the documentation of results in the secondary articles, these mistakes will be reflected in this study as well.

6 Conclusion and future work

This study provides the results of a tertiary study on sentiment analysis methods whereby we aimed to highlight the adopted features (input/output), adopted approaches, the adopted data sets, and the challenges with respect to sentiment analysis. The answers to the research questions were derived based on in-depth secondary studies.

A different number of input and output features could be identified. Interestingly, some features appeared to be described in all the secondary studies, while other features were more specific to a selected set of secondary studies. The results further indicate that sentiment analysis has been applied in various domains, among which social media is the most popular. Also, the study showed that different domains require the use of different techniques.

There also seems a trend towards using more complex deep learning techniques, since they can detect more complex patterns in text and perform particularly well with larger datasets. In some use cases like, for example, advertisement, slight improvements in performance that can be obtained through deep learning can have a great impact. However, it should be noted that traditional machine learning models are less computationally expensive and perform sufficiently well for sentiment analysis tasks. They are widely praised for their performance and efficiency.

This study showed that the most prominent challenges in sentiment analysis are domain and language dependency. Specific text corpora are required for differing languages and domains of interest. Attempts for cross-domain and multi-lingual sentiment analysis models have been made, but this challenging task should be explored further. Other prominent challenges are opinion spam detection and the application of deep learning for sentiment analysis tasks. Overall, the study shows that sentiment analysis is a timely and important

research topic. The adoption of a tertiary study showed the additional value that could not be derived from each of the secondary studies.

The following future directions and challenges have also been mainly discussed in deep learning-based survey papers: New datasets are required for more challenging tasks, common sense knowledge must be modeled, interpretable deep learning-based models must be developed, and memory-efficient models are required (Minaee et al. 2020). Domain adaptation techniques are needed, multi-lingual applications should be addressed, technical requirements such as a huge amount of labeled data requirement must be considered, and linguistic complications must be investigated (Do et al. 2019). Popular deep learning techniques such as deep reinforcement learning and generative adversarial networks can be evaluated to solve some challenging tasks, advantages of the BERT algorithm can be considered, language structures (e.g., slangs) can be investigated in detail, dynamic sentiment analysis can be studied, and sentiment analysis for heterogeneous data can be implemented (Habimana et al. 2020a). Dependency trees in recursive neural networks can be investigated, domain adaptation can be analyzed in detail, and linguistic-subjective phenomena (e.g., irony and sarcasm) can be studied (Rojas-Barahona 2016). Different applications of sentiment analysis (e.g., medical domain and security screening of employees) can be implemented, and transfer learning approaches can be analyzed for sentiment classification (Yadav and Vishwakarma 2019). Comparative studies should be extended with new approaches and new datasets, and also hybrid approaches to reduce computational cost and improve performance must be developed (Dang et al. 2020).

Appendix

See Table 16.

Table 16 Recent state-of-the-art deep learning-based sentiment analysis papers

ID	References	Title	Deep learning algorithms used
P1	Ray and Chakrabarti (2020)	A mixed approach of deep learning method and rule-based method to improve aspect level sentiment analysis	CNN
P2	Kumar et al. (2020a)	Hybrid context enriched deep learning model for fine-grained sentiment analysis in textual and visual semiotic modality social data	CNN
P3	Sankar et al. (2020)	Intelligent sentiment analysis approach using edge computing-based deep learning technique	CNN
P4	Yang et al. (2020)	Sentiment analysis for E-commerce product reviews in chinese based on sentiment lexicon and deep learning	CNN, GRU
P5	Li et al. (2020a)	How textual quality of online reviews affect classification performance: A case of deep learning sentiment analysis	CNN, RNN, LSTM
P6	Colón-Ruiz and Segura-Bedmar (2020)	Comparing deep learning architectures for sentiment analysis on drug reviews	BERT, CNN, LSTM
P7	Ghorbani et al. (2020)	ConvLSTMConv network: a deep learning approach for sentiment analysis in cloud computing	CNN, LSTM
P8	Liu (2020)	Text sentiment analysis based on CBOW model and deep learning in big data environment	CNN
P9	Li et al. (2020b)	HEMOS: A novel deep learning-based fine-grained humor detecting method for sentiment analysis of social media	LSTM
P10	Studiawan et al. (2020)	Sentiment analysis in a forensic timeline with deep learning	RNN, GRU, LSTM
P11	Sadr et al. (2020)	Multi-view deep network: A deep model based on learning features from heterogeneous neural networks for sentiment analysis	CNN, Recursive Neural Network (ReNN)
P12	Yadav and Vishwakarma (2020a, b)	A deep learning architecture of RA-DLNet for visual sentiment analysis	CNN
P13	Ombabi et al. (2020)	Deep learning CNN-LSTM framework for Arabic sentiment analysis using textual information shared in social networks	CNN, LSTM
P14	Hung (2020a)	Domain-Specific versus general-purpose word representations in sentiment analysis for deep learning models	CNN, LSTM
P15	Besseio and Elmousalami (2020)	Subword attentive model for Arabic sentiment analysis: A deep learning approach	CNN, RNN

Table 16 (continued)

ID	References	Title	Deep learning algorithms used
P16	Onan (2020a)	Sentiment analysis on massive open online course evaluations: A text mining and deep learning approach	CNN, RNN, LSTM, GRU
P17	Khedkar and Shinde (2020a)	Deep learning-based approach to classify praises or complaints	CNN
P18	Seo et al. (2020)	Comparative study of deep learning-based sentiment classification	CNN, RNN
P19	Wadawadagi and Pagi (2020)	Sentiment analysis with deep neural networks: comparative study and performance assessment	CNN, LSTM, GRU, RNN, ReNN
P20	Salur and Aydin (2020)	A novel hybrid deep learning model for sentiment classification	CNN, LSTM, GRU
P21	Aslam et al. (2020)	A novel framework for sentiment analysis using deep learning	CNN, LSTM, RNN
P22	Naseem et al. (2020)	Transformer based deep intelligent contextual embedding for twitter sentiment analysis	LSTM
P23	Onan (2020b)	Sentiment analysis on product reviews based on weighted word embeddings and deep neural networks. Concurrency and Computation: Practice and Experience, e5909	CNN, LSTM
P24	Krouska et al. (2020)	Deep learning for Twitter sentiment analysis: The effect of pre-trained word embedding	CNN
P25	Luo et al. (2020)	A fine-grained sentiment analysis of online guest reviews of economy hotels in China	LSTM
P26	Haralabopoulos et al. (2020)	Ensemble deep learning for multilabel binary classification of user-generated content	CNN, LSTM, DNN
P27	Chandra and Jana (2020)	Sentiment analysis using machine learning and deep learning	CNN, LSTM
P28	Dessí et al. (2020)	Deep learning adaptation with word embeddings for sentiment analysis on online course reviews	LSTM
P29	Dashtipour et al. (2020)	A hybrid Persian sentiment analysis framework: Integrating dependency grammar based rules and deep neural networks	CNN, LSTM
P30	Sun and He (2020)	A novel approach to generate a large scale of supervised data for short text sentiment analysis	CNN, LSTM, GAN
P31	Kumar and Sharan (2020)	Deep learning-based frameworks for aspect-based sentiment analysis	CNN, LSTM

Table 16 (continued)

ID	References	Title	Deep learning algorithms used
P32	Kansara and Sawant (2020)	Comparison of traditional machine learning and deep learning approaches for sentiment analysis	CNN, LSTM
P33	Lim et al. (2020)	Tweet sentiment analysis using deep learning with nearby locations as features	CNN
P34	Che and Li (2020)	HCI with DEEP learning for sentiment analysis of corporate social responsibility report	ReNN
P35	Tao and Fang (2020)	Toward multi-label sentiment analysis: a transfer learning based approach	BERT, CNN, LSTM
P36	Kabra and Shrawne (2020)	Location-wise news headlines classification and sentiment analysis: A deep learning approach	LSTM
P37	Rao and Ranjana (2020)	Deep learning method to identify the demographic attribute to enhance effectiveness of sentiment analysis	DNN
P38	Venkatakrishnan et al. (2020)	Sentiment analysis on Google Play Store data using deep learning	DNN
P39	Khan and Malviya (2020)	Big data approach for sentiment analysis of twitter data using Hadoop framework and deep learning	RNN
P40	Ullah et al. (2020)	An algorithm and method for sentiment analysis using the text and emoticon	CNN, LSTM
P41	Kumar et al. (2020b)	Aspect-based sentiment analysis using deep networks and stochastic optimization	CNN
P42	Park et al. (2020)	Deep learning models and datasets for aspect term sentiment classification: Implementing holistic recurrent attention on target-dependent memories	LSTM, ReNN
P43	Basiri et al. (2020)	A novel method for sentiment classification of drug reviews using fusion of deep and machine learning techniques	CNN, GRU
P44	Kumar and Garg (2020)	Aspect-based sentiment analysis using deep learning convolutional neural network	CNN
P45	Valdivia et al. (2020)	What do people think about this monument? Understanding negative reviews via deep learning, clustering and descriptive rules	CNN

Table 16 (continued)

ID	References	Title	Deep learning algorithms used
P46	Da'u et al. (2020a)	Recommendation system exploiting aspect-based opinion mining with deep learning method	CNN
P47	Kula et al. (2020)	Sentiment analysis for fake news detection by means of neural networks	RNN, LSTM
P48	Hameed and Garcia-Zapirain (2020)	Sentiment classification using a single-layered BiLSTM model	GRU, CNN, LSTM
P49	Alam et al. (2020)	Social media sentiment analysis through parallel dilated convolutional neural network for smart city applications	CNN
P50	Singh et al. (2020)	Identifying hidden sentiment in text using deep neural network	LSTM, CNN
P51	Giménez et al. (2020)	Semantic-based padding in convolutional neural networks for improving the performance in natural language processing. A case of study in sentiment analysis	CNN
P52	Patel et al. (2020)	Sentiment analysis on movie review using deep learning RNN method	RNN
P53	Ren et al. (2020a)	A Lexicon-Enhanced Attention Network for Aspect-Level Sentiment Analysis	CNN, LSTM
P54	Li et al. (2020c)	User reviews: Sentiment analysis using lexicon integrated two-channel CNN-LSTM family models	CNN, LSTM
P55	Alexandridis et al. (2020)	A Deep learning approach to aspect-based sentiment prediction	LSTM
P56	Khedkar and Shinde (2020b)	Deep learning and ensemble approach for praise or complaint classification	CNN
P57	Huang et al. (2020a, b, c)	Attention-based modality-gated networks for image-text sentiment analysis	CNN, LSTM
P58	Cai et al. (2020)	Recurrent neural network with pooling operation and attention mechanism for sentiment analysis: A multi-task learning approach	CNN, RNN, LSTM
P59	Da'u et al. (2020b)	Weighted aspect-based opinion mining using deep learning for recommender system	CNN

Table 16 (continued)

ID	References	Title	Deep learning algorithms used
P60	Tran et al. (2020)	Bidirectional independently long short-term memory and conditional random field integrated model for aspect extraction in sentiment analysis	LSTM
P61	Gan et al. (2020a)	Sparse attention based separable dilated convolutional neural network for targeted sentiment analysis	CNN, LSTM
P62	Usama et al. (2020)	Attention-based sentiment analysis using convolutional and recurrent neural network	CNN, RNN, LSTM
P63	Parimala et al. (2020)	Spatiotemporal-based sentiment analysis on tweets for risk assessment of event using deep learning approach	CNN, LSTM
P64	Hung (2020b)	Integrating Sentiment Analysis in Recommender Systems	CNN, LSTM
P65	Liu and Shen (2020a)	Aspect-based sentiment analysis with gated alternate neural network	Gated Alternate Neural Network (GANN)
P66	Yildirim (2020)	Comparing deep neural networks to traditional models for sentiment analysis in Turkish language	RNN, LSTM, GRU
P67	Dong et al. (2020a)	Variable convolution and pooling convolutional neural network for text sentiment classification	CNN, LSTM
P68	Sangeetha and Prabha (2020)	Sentiment analysis of student feedback using multi-head attention fusion model of word and context embedding for LSTM	LSTM
P69	Kiran et al. (2020)	OSLFFit (Organic Simultaneous LSTM and CNN Fit): A novel deep learning based solution for sentiment polarity classification of reviews	CNN, LSTM
P70	Shakeel and Karim (2020)	Adapting deep learning for sentiment classification of code-switched informal short text	CNN, LSTM
P71	Su et al. (2020)	A novel LMAEB-CNN model for Chinese microblog sentiment analysis	CNN, LSTM
P72	Jia et al. (2020)	Hierarchical gated deep memory network with position-aware for aspect-based sentiment analysis	LSTM, GRU

Table 16 (continued)

ID	References	Title	Deep learning algorithms used
P73	Ren et al. (2020b)	DNnet: A lightweight and efficient model for aspect based sentiment analysis	CNN, LSTM, BERT, Distillation Network (DN)
P74	Xi et al. (2020)	Domain adaptation with category attention network for deep sentiment analysis	CNN, Category Attention Network (CAN)
P75	Onan (2020c)	Mining opinions from instructor evaluation reviews: A deep learning approach	CNN, RNN, GRU, LSTM
P76	Jin et al. (2020)	Multi-task learning model based on multi-scale CNN and LSTM for sentiment classification	CNN, LSTM
P77	Han et al. (2020a, b)	Aspect-level drug reviews sentiment analysis based on double BiGRU and knowledge transfer	CNN, LSTM, GRU
P78	Liu and Shen (2020b)	ReMemNN: A novel memory neural network for powerful interaction in aspect-based sentiment analysis	LSTM, Recurrent Memory Neural Network (ReMemNN)
P79	Lu et al. (2020a)	Interactive rule attention network for aspect-level sentiment analysis	LSTM, BERT, IRAN
P80	Shuang et al. (2020)	Feature distillation network for aspect-based sentiment analysis	LSTM, Feature Distillation Network (FDN)
P81	Meškela and Frasincar (2020)	ALDONAR: A hybrid solution for sentence-level aspect-based sentiment analysis using a lexicalized domain ontology and a regularized neural attention model	CNN, GRU, LSTM BERT
P82	Gan et al. (2020b)	Multi-entity sentiment analysis using self-attention based hierarchical dilated convolutional neural network	LSTM, Self-Attention based Hierarchical Dilated Convolutional Neural Network (SA-HDCNN)
P83	Ren et al. (2020c)	Sarcasm detection with sentiment semantics enhanced multi-level memory network	CNN, LSTM, RNN, DNN
P84	Lin et al. (2020)	Sentiment analysis with comparison enhanced deep neural network	CNN, LSTM, RNN
P85	Ling et al. (2020)	Hybrid neural network for Sina Weibo sentiment analysis	CNN, LSTM
P86	Zhang et al. (2020a)	Knowledge guided capsule attention network for aspect-based sentiment analysis	LSTM, CapsN

Table 16 (continued)

ID	References	Title	Deep learning algorithms used
P87	Yao and Wang (2020)	Domain-specific sentiment analysis for tweets during hurricanes (DSSA-H): A domain-adversarial neural-network-based approach	LSTM, GRU
P88	Ji and Wu (2020)	Cascade architecture with rhetoric long short-term memory for complex sentence sentiment analysis	LSTM, BERT
P89	Abid et al. (2020)	Multi-source social media data sentiment analysis using bidirectional recurrent convolutional neural networks	CNN, GRU, RNN, LSTM
P90	Jiang et al. (2020)	Fusion-extraction network for multimodal sentiment analysis	CNN, LSTM, Fusion-Extraction Network (FENet)
P91	Lou et al. (2020)	Emoji-based sentiment analysis using attention networks	LSTM, GRU
P92	Wang et al. (2020)	Emotion-semantic-enhanced bidirectional LSTM with multi-head attention mechanism for microblog sentiment analysis	CNN, LSTM
P93	Pan and Liang (2020)	Chinese text sentiment analysis based on Bi-GRU and self-attention	GRU, LSTM
P94	Zhu et al. (2020)	Attention based BiLSTM-MCNN for sentiment analysis	CNN, LSTM
P95	Pröllochs et al. (2020)	Negation scope detection for sentiment analysis: A reinforcement learning framework for replicating human interpretations	Deep Q-Network
P96	Peng et al. (2020)	Knowing what, how and why: A near complete solution for aspect-based sentiment analysis	LSTM
P97	Liu and Mukaidani (2020)	Effective-target representation via LSTM with attention for aspect-level sentiment analysis	LSTM
P98	Huang et al. (2020a, b, c)	Sentiment strength detection with a context-dependent lexicon-based convolutional neural network	CNN, LSTM
P99	Zao et al. (2020)	Context-specific heterogeneous graph convolutional network for implicit sentiment analysis	LSTM, Graph Convolutional Neural Network (GCN)
P100	Cheng et al. (2020)	Text sentiment orientation analysis based on multi-channel CNN and bidirectional GRU with attention mechanism	CNN, GRU, LSTM
P101	Zhou et al. (2020)	Is position important? deep multi-task learning for aspect-based sentiment analysis	LSTM

Table 16 (continued)

ID	References	Title	Deep learning algorithms used
P102	Huang et al. (2020c)	Lexicon-based sentiment convolutional neural networks for online review analysis	CNN, LSTM
P103	Mahmood et al. (2020)	Deep sentiments in Roman Urdu text using Recurrent Convolutional Neural Network model	Recurrent Convolutional Neural Network (RCNN)
P104	Kumar and Malarvizhi (2020)	Bi-directional LSTM-CNN combined method for sentiment analysis in part of speech tagging (PoS)	CNN, LSTM
P105	Habimana et al. (2020b)	Attentive convolutional gated recurrent network: a contextual model to sentiment analysis	CNN, RNN, GRU, RCNN, LSTM
P106	Alarifi et al. (2020)	A big data approach to sentiment analysis using greedy feature selection with cat swarm optimization-based long short-term memory neural networks	LSTM
P107	Gu et al. (2020)	Sentiment analysis via deep multichannel neural networks with variational information bottleneck	CNN, GRU, LSTM
P108	Dong et al. (2020b)	A sentiment analysis method of capsule network based on BiLSTM	CNN, LSTM, CapsN
P109	Lu et al. (2020b)	Multi-task learning using variational auto-encoder for sentiment classification	CNN, LSTM, Autoencoder
P110	Zhao et al. (2020)	Modeling sentiment dependencies with graph convolutional networks for aspect-level sentiment classification	GRU, LSTM, BERT, GCN
P111	Zhang et al. (2020b)	Multi-layer attention based CNN for target-dependent sentiment classification	CNN, LSTM
P112	Lai et al. (2020)	Fine-grained emotion classification of Chinese microblogs based on graph convolution networks	CNN, LSTM, GCN

Funding Open Access funding provided by the Qatar National Library.

Compliance with Ethical Standards

Conflict of interest The authors have no conflicts of interest to declare that are relevant to the content of this article.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Abid F, Li C, Alam M (2020) Multi-source social media data sentiment analysis using bidirectional recurrent convolutional neural networks. *Comput Commun* 157:102–115
- Ahmad M, Aftab S, Ali I, Hameed N (2017) Hybrid tools and techniques for sentiment analysis: a review 8(4):7
- Ahmad M, Aftab S, Bashir MS, Hameed N (2018) Sentiment analysis using SVM: a systematic literature review. *Int J Adv Comput Sci Appl* 9(2):182–188 (**Scopus**)
- Ahmed Ibrahim M, Salim N (2013) Opinion analysis for twitter and Arabic tweets: a systematic literature review. *J Theor Appl Inf Technol* 56(3):338–348 (**Scopus**)
- Alam M, Abid F, Guangpei C, Yunrong LV (2020) Social media sentiment analysis through parallel dilated convolutional neural network for smart city applications. *Comput Commun* 154:129–137
- Alarifi A, Tolba A, Al-Makhadmeh Z, Said W (2020) A big data approach to sentiment analysis using greedy feature selection with cat swarm optimization-based long short-term memory neural networks. *J Supercomput* 76(6):4414–4429
- Alexandridis G, Michalakakis K, Aliprantis J, Polydoros P, Tsantilas P, Caridakis G (2020) A deep learning approach to aspect-based sentiment prediction. In: IFIP International conference on artificial intelligence applications and innovations. Springer, Cham, pp 397–408
- Al-Moslmi T, Omar N, Abdullah S, Albared M (2017) Approaches to cross-domain sentiment analysis: a systematic literature review. *IEEE Access* 5:16173–16192 (**Scopus**)
- Almotairi, M. (2009) A framework for successful CRM implementation. In: European and mediterranean conference on information systems. pp 1–14
- Aslam A, Qamar U, Saqib P, Ayesha R, Qadeer A (2020) A novel framework for sentiment analysis using deep learning. In: 2020 22nd International conference on advanced communication technology (ICACT). IEEE, pp 525–529
- Basiri ME, Abdar M, Cifci MA, Nemati S, Acharya UR (2020) A novel method for sentiment classification of drug reviews using fusion of deep and machine learning techniques. *Knowl-Based Syst* 198:1–19
- Becker JU, Greve G, Albers S (2009) The impact of technological and organizational implementation of CRM on customer acquisition, maintenance, and retention. *Int J Res Mark* 26(3):207–215
- Behera RN, Manan R, Dash S (2016) Ensemble based hybrid machine learning approach for sentiment classification-a review. *Int J Comput Appl* 146(6):31–36
- Beseiso M, Elmousalami H (2020) Subword attentive model for arabic sentiment analysis: a deep learning approach. *ACM Trans Asian Low-Resour Lang Inf Process (TALLIP)* 19(2):1–17
- Blum A, Mitchell T (1998) Combining labeled and unlabeled data with co-training. *Proc Eleventh Ann Conf Comput Learn Theory COLT* 98:92–100
- Bondielli A, Marcelloni F (2019) A survey on fake news and rumour detection techniques. *Inf Sci* 497:38–55
- Budgen D, Brereton P, Drummond S, Williams N (2018) Reporting systematic reviews: some lessons from a tertiary study. *Inf Softw Technol* 95:62–74
- Cadavid H, Andrikopoulos V, Avgeriou P (2020) Architecting systems of systems: a tertiary study. *Inf Softw Technol* 118(106202):1–18

- Cai Y, Huang Q, Lin Z, Xu J, Chen Z, Li Q (2020) Recurrent neural network with pooling operation and attention mechanism for sentiment analysis: a multi-task learning approach. *Knowl-Based Syst* 203(105856):1–12
- Cambria E (2016) Affective computing and sentiment analysis. *IEEE Intell Syst* 31(2):102–107
- Cambria E, Hussain A, Havasi C, Eckl C (2009) Common sense computing: from the society of mind to digital intuition and beyond. In: European workshop on biometrics and identity management. Springer, Berlin, Heidelberg, pp 252–259
- Cambria E, Poria S, Gelbukh A, Thelwall M (2017) Sentiment analysis is a big suitcase. *IEEE Intell Syst* 32(6):74–80
- Cambria E, Li Y, Xing FZ, Poria S, Kwok K (2020) SenticNet 6: ensemble application of symbolic and subsymbolic AI for sentiment analysis. In: Proceedings of the 29th ACM International conference on information & knowledge management. pp 105–114
- Can EF, Ezen-Can A., & Can, F. (2018). Multi-lingual sentiment analysis: an RNN-based framework for limited data. In: Proceedings of ACM SIGIR 2018 workshop on learning from limited or noisy data, Ann Arbor
- Catal C, Mishra D (2013) Test case prioritization: a systematic mapping study. *Softw Qual J* 21(3):445–478
- Chandra Y, Jana A (2020) Sentiment analysis using machine learning and deep learning. In: 2020 7th International conference on computing for sustainable global development (INDIACom). IEEE, pp. 1–4
- Chapelle O, Zien A (2005) Semi-supervised classification by low density separation. In: AISTATS vol 2005, pp 57–64
- Che S, Li X (2020) HCI with DEEP learning for sentiment analysis of corporate social responsibility report. *Curr Psychol*. <https://doi.org/10.1007/s12144-020-00789-y>
- Chen JJ, Popovich K (2003) Understanding customer relationship management (CRM). *Bus Process Manag* 9(5):672–688. <https://doi.org/10.1108/14637150310496758>
- Chen L, Chen G, Wang F (2015) Recommender systems based on user reviews: the state of the art. *User Model User-Adap Inter* 25(2):99–154. <https://doi.org/10.1007/s11257-015-9155-5>
- Chen H, Sun M, Tu C, Lin Y, Liu Z (2016) Neural sentiment classification with user and product attention. In: Proceedings of the 2016 conference on empirical methods in natural language processing. pp 1650–1659
- Chen P, Sun Z, Bing L, Yang W (2017) Recurrent attention network on memory for aspect sentiment analysis. In: Proceedings of the 2017 conference on empirical methods in natural language processing. pp 452–461
- Chen H, Liu J, Lv Y, Li MH, Liu M, Zheng Q (2018) Semi-supervised clue fusion for spammer detection in Sina Weibo. *Inf Fusion* 44:22–32. <https://doi.org/10.1016/j.inffus.2017.11.002>
- Cheng Y, Yao L, Xiang G, Zhang G, Tang T, Zhong L (2020) Text sentiment orientation analysis based on multi-channel CNN and bidirectional GRU with attention mechanism. *IEEE Access* 8:134964–134975
- Choi Y, Cardie C (2008) Learning with compositional semantics as structural inference for subsentential sentiment analysis. In: Proceedings of the 2008 conference on empirical methods in natural language processing. pp 793–801
- Colón-Ruiz C, Segura-Bedmar I (2020) Comparing deep learning architectures for sentiment analysis on drug reviews. *J Biomed Inform* 110(103539):1–11
- Crawford M, Khoshgoftaar TM, Prusa JD, Richter AN, Al Najada H (2015) Survey of review spam detection using machine learning techniques. *J Big Data* 2(1):23. <https://doi.org/10.1186/s40537-015-0029-9>
- Cruzes DS, Dybå T (2011) Research synthesis in software engineering: a tertiary study. *Inf Softw Technol* 53(5):440–455
- Curcio K, Santana R, Reinehr S, Malucelli A (2019) Usability in agile software development: a tertiary study. *Comput Stand Interfaces* 64:61–77. <https://doi.org/10.1016/j.csi.2018.12.003>
- Da Silva NFF, Coletta LFS, Hruschka ER, Hruschka ER Jr (2016a) Using unsupervised information to improve semi-supervised tweet sentiment classification. *Inf Sci* 355:348–365. <https://doi.org/10.1016/j.ins.2016.02.002>
- Dang NC, Moreno-García MN, De la Prieta F (2020) Sentiment analysis based on deep learning: a comparative study. *Electronics* 9(3):483
- Dashtipour K, Gogate M, Li J, Jiang F, Kong B, Hussain A (2020) A hybrid persian sentiment analysis framework: Integrating dependency grammar based rules and deep neural networks. *Neurocomputing* 380:1–10
- Da’u A, Salim N, Rabiu I, Osman A (2020a) Recommendation system exploiting aspect-based opinion mining with deep learning method. *Inf Sci* 512:1279–1292
- Da’u A, Salim N, Rabiu I, Osman A (2020b) Weighted aspect-based opinion mining using deep learning for recommender system. *Expert Syst Appl* 140(112871):1–12

- De Oliveira Lima T, Colaco Junior M, Nunes MASN (2018) Mining on line general opinions about sustainability of hotels: a systematic literature mapping. In: Gervasi O, Murgante B, Misra S, Stankova E, Torre CM, Rocha AMAC, Taniar D, Apduhan BO, Tarantino E, Ryu Y (eds) Computational science and its applications–ICCSA 2018. Springer, New York, pp 558–574
- Dessí D, Dragoni M, Fenu G, Marras M, Recupero DR (2020) Deep learning adaptation with word embeddings for sentiment analysis on online course reviews. deep learning-based approaches for sentiment analysis. Springer, Singapore, pp 57–83
- Devlin J, Chang MW, Lee K, Toutanova K (2019) BERT: Pre-training of deep bidirectional transformers for language understanding. In: Proceedings of the 2019 conference of the north american chapter of the association for computational linguistics: human language technologies vol 1 (Long and Short Papers). pp 4171–4186
- Dietterich TG (2002) Machine learning for sequential data: a review. In: Caelli T, Amin A, Duin RPW, de Ridder D, Kamel M (eds) Structural, syntactic, and statistical pattern recognition. Springer, Berlin Heidelberg, pp 15–30
- Do HH, Prasad PWC, Maag A, Alsadoon A (2019) Deep learning for aspect-based sentiment analysis: a comparative review. *Expert Syst Appl* 118:272–299
- Dong M, Li Y, Tang X, Xu J, Bi S, Cai Y (2020a) Variable convolution and pooling convolutional neural network for text sentiment classification. *IEEE Access* 8:16174–16186
- Dong Y, Fu Y, Wang L, Chen Y, Dong Y, Li J (2020b) A sentiment analysis method of capsule network based on BiLSTM. *IEEE Access* 8:37014–37020
- Duan J, Luo B, Zeng J (2020) Semi-supervised Learning with generative model for sentiment classification of stock messages. *Expert Syst Appl* 158(113540):1–9
- Ebrahimi M, Yazdavar AH, Sheth A (2017) Challenges of sentiment analysis for dynamic events. *IEEE Intell Syst* 32(5):70–75
- Elmuti D, Jia H, Gray D (2009) Customer relationship management strategic application and organizational effectiveness: An empirical investigation. *J Strateg Mark* 17(1):75–96. <https://doi.org/10.1080/09652540802619301>
- Filatova, E. (2012). Irony and sarcasm: corpus generation and analysis using crowdsourcing. In: *Lrec*, pp 392–398
- Gan C, Wang L, Zhang Z, Wang Z (2020a) Sparse attention based separable dilated convolutional neural network for targeted sentiment analysis. *Knowl-Based Syst* 188(104827):1–10
- Gan C, Wang L, Zhang Z (2020b) Multi-entity sentiment analysis using self-attention based hierarchical dilated convolutional neural network. *Future Gener Comput Syst* 112:116–125
- Genc-Nayebi N, Abran A (2017) A systematic literature review: opinion mining studies from mobile app store user reviews. *J Syst Softw* 125:207–219. <https://doi.org/10.1016/j.jss.2016.11.027>
- Ghorbani M, Bahaghighat M, Xin Q, Özen F (2020) ConvLSTMConv network: a deep learning approach for sentiment analysis in cloud computing. *J Cloud Comput* 9(1):1–12
- Gieseke F, Airola A, pahikkala T, Oliver K (2012) Sparse quasi-newton optimization for semi-supervised support vector machines. Proceedings of the 1st international conference on pattern recognition applications and methods 45–54. <https://doi.org/10.5220/0003755300450054>
- Giménez M, Palanca J, Botti V (2020) Semantic-based padding in convolutional neural networks for improving the performance in natural language processing. a case of study in sentiment analysis. *Neurocomputing* 378:315–323
- Gneiser MS (2010) Value-Based CRM. *Bus Inf Syst Eng* 2(2):95–103. <https://doi.org/10.1007/s12599-010-0095-7>
- Goldberg AB, Zhu X (2006) Seeing stars when there aren't many stars: graph-based semi-supervised learning for sentiment categorization. In: Proceedings of textgraphs: the first workshop on graph based methods for natural language processing. pp 45–52
- Goulão M, Amaral V, Mernik M (2016) Quality in model-driven engineering: a tertiary study. *Softw Qual J* 24(3):601–633. <https://doi.org/10.1007/s11219-016-9324-8>
- Gu T, Xu G, Luo J (2020) Sentiment analysis via deep multichannel neural networks with variational information bottleneck. *IEEE Access* 8:121014–121021
- Gupta R, Sahu S, Espy-Wilson C, Narayanan S (2018) Semi-supervised and transfer learning approaches for low resource sentiment classification. In: 2018 IEEE International conference on acoustics, speech and signal processing (ICASSP). IEEE, pp 5109–5113
- Habimana O, Li Y, Li R, Gu X, Yu G (2020a) Sentiment analysis using deep learning approaches: an overview. *Sci China Inf Sci* 63(1):1–36
- Habimana O, Li Y, Li R, Gu X, Yan W (2020b) Attentive convolutional gated recurrent network: a contextual model to sentiment analysis. *Int J Mach Learn Cybern* 11:2637–2651

- Hameed Z, Garcia-Zapirain B (2020) Sentiment classification using a single-layered BiLSTM model. *IEEE Access* 8:73992–74001
- Han Y, Liu Y, Jin Z (2020a) Sentiment analysis via semi-supervised learning: a model based on dynamic threshold and multi-classifiers. *Neural Comput Appl* 32(9):5117–5129
- Han Y, Liu M, Jing W (2020b) Aspect-level drug reviews sentiment analysis based on double BiGRU and knowledge transfer. *IEEE Access* 8:21314–21325
- Haralabopoulos G, Anagnostopoulos I, McAuley D (2020) Ensemble deep learning for multilabel binary classification of user-generated content. *Algorithms* 13(4):83
- Hassan R, Islam MR (2019) Detection of fake online reviews using semi-supervised and supervised learning. In: 2019 International conference on electrical, computer and communication engineering (ECCE). pp 1–5. <https://doi.org/https://doi.org/10.1109/ECACE.2019.8679186>
- Hemmatian F, Sohrabi MK (2017) A survey on classification techniques for opinion mining and sentiment analysis. *Artif Intell Rev* 52(3):1495–1545. <https://doi.org/10.1007/s10462-017-9599-6>
- Huang M, Xie H, Rao Y, Feng J, Wang FL (2020b) Sentiment strength detection with a context-dependent lexicon-based convolutional neural network. *Inf Sci* 520:389–399
- Huang F, Wei K, Weng J, Li Z (2020a) Attention-based modality-gated networks for image-text sentiment analysis. *ACM Trans Multimed Comput Commun Appl (TOMM)* 16(3):1–19
- Huang M, Xie H, Rao Y, Liu Y, Poon LK, Wang FL (2020c) Lexicon-based sentiment convolutional neural networks for online review analysis. *IEEE Trans Affect Comput* (Early Access), 1–1
- Hung BT (2020) Domain-specific versus general-purpose word representations in sentiment analysis for deep learning models. *Frontiers in intelligent computing: theory and applications*. Springer, Singapore, pp 252–264
- Hung BT (2020) Integrating sentiment analysis in recommender systems. *Reliability and statistical computing*. Springer, Cham, pp 127–137
- Hussain A, Cambria E (2018) Semi-supervised learning for big social data analysis. *Neurocomputing* 275:1662–1673
- Ishaya T, Folarin M (2012) A service oriented approach to business intelligence in telecoms industry. *Telemat Inform* 29(3):273–285. <https://doi.org/10.1016/j.tele.2012.01.004>
- Ji C, Wu H (2020) Cascade architecture with rhetoric long short-term memory for complex sentence sentiment analysis. *Neurocomputing* 405:161–172
- Jia Z, Bai X, Pang S (2020) Hierarchical gated deep memory network with position-aware for aspect-based sentiment analysis. *IEEE Access* 8:136340–136347
- Jiang T, Wang J, Liu Z, Ling Y (2020) Fusion-extraction network for multimodal sentiment analysis. *Pacific-Asia conference on knowledge discovery and data mining*. Springer, Cham, pp 785–797
- Jin N, Wu J, Ma X, Yan K, Mo Y (2020) Multi-task learning model based on multi-scale CNN and LSTM for sentiment classification. *IEEE Access* 8:77060–77072
- Josiassen A, Assaf AG, Cvelbar LK (2014) CRM and the bottom line: do all CRM dimensions affect firm performance? *Int J Hosp Manag* 36:130–136. <https://doi.org/10.1016/j.ijhm.2013.08.005>
- Kabra A, Shrawne S (2020) Location-wise news headlines classification and sentiment analysis: a deep learning approach. *International conference on intelligent computing and smart communication 2019*. Springer, Singapore, pp 383–391
- Kamal A (2013) Subjectivity classification using machine learning techniques for mining feature-opinion pairs from web opinion sources 10(5):191–200
- Kamal N, Andrew M, Tom M (2006) Semi-supervised text classification using EM. In: Chapelle O, Scholkopf B, Zien A (eds) *Semi-supervised learning*. The MIT Press, Cambridge, pp 32–55. <https://doi.org/10.7551/mitpress/9780262033589.003.0003>
- Kansara D, Sawant V (2020) Comparison of traditional machine learning and deep learning approaches for sentiment analysis. *Advanced computing technologies and applications*. Springer, Singapore, pp 365–377
- Karimpour J, Noroozi AA, Alizadeh S (2012) Web spam detection by learning from small labeled samples. *Int J Comput Appl* 50(21):1–5. <https://doi.org/10.5120/7924-0993>
- Kasmuri E, Basiron H (2017) Subjectivity analysis in opinion mining—a systematic literature review. *Int J Adv Soft Comput Appl* 9(3):132–159 (**Scopus**)
- Khan M, Malviya A (2020) Big data approach for sentiment analysis of twitter data using Hadoop framework and deep learning. In: 2020 International conference on emerging trends in information technology and engineering (ic-ETITE). *IEEE*, pp 1–5
- Khedkar S, Shinde S (2020) Deep learning and ensemble approach for praise or complaint classification. *Procedia Comput Sci* 167:449–458

- Khedkar S, Shinde S (2020a) Deep learning-based approach to classify praises or complaints. In: Proceedings of international conference on computational science and applications: ICCSA 2019. Springer, New York, p 391
- Kim Y (2014) Convolutional neural networks for sentence classification. In: Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP). pp 1746–1751
- Kim H-S, Kim Y-G (2009) A CRM performance measurement framework: Its development process and application. *Ind Mark Manag* 38(4):477–489. <https://doi.org/10.1016/j.indmarman.2008.04.008>
- Kiran R, Kumar P, Bhasker B (2020) OSLCFit (organic simultaneous LSTM and CNN fit): a novel deep learning based solution for sentiment polarity classification of reviews. *Expert Syst Appl* 157(113488):1–12
- Kitchenham B (2004) Procedures for performing systematic reviews, vol 33. Keele University, Keele, UK, pp 1–26
- Kitchenham B, Charters S (2007) Guidelines for performing systematic literature reviews in software engineering version 2.3. *Engineering*, 45(4ve), p 1051
- Kitchenham BA, Dyba T, Jorgensen M (2004) Evidence-based software engineering. In: Proceedings 26th international conference on software engineering. IEEE, pp 273–281
- Kitchenham B, Pretorius R, Budgen D, Pearl Brereton O, Turner M, Niazi M, Linkman S (2010) Systematic literature reviews in software engineering—a tertiary study. *Inf Softw Technol* 52(8):792–805. <https://doi.org/10.1016/j.infsof.2010.03.006>
- Kitchenham BA, Budgen D, Brereton OP (2010b) The value of mapping studies—a participant-observer case study. In: 14th international conference on evaluation and assessment in software engineering (ease). pp 1–9
- Kitchenham BA, Budgen D, Pearl Brereton O (2011) Using mapping studies as the basis for further research—a participant-observer case study. *Inf Softw Technol* 53(6):638–651. <https://doi.org/10.1016/j.infsof.2010.12.011>
- Koksal O, Tekinerdogan B (2017) Feature-driven domain analysis of session layer protocols of internet of things. *IEEE Int Congr Internet Things (ICIOT) 2017*:105–112. <https://doi.org/10.1109/IEEE.ICIOT.2017.19>
- Krouska A, Troussas C, Virvou M (2020) Deep learning for twitter sentiment analysis: the effect of pre-trained word embedding. *Machine learning paradigms*. Springer, Cham, pp 111–124
- Kula S, Choraś M, Kozik R, Ksieniewicz P, Woźniak M (2020) Sentiment analysis for fake news detection by means of neural networks. *International conference on computational science*. Springer, Cham, pp 653–666
- Kumar V (2010) Customer relationship management. In: *Wiley international encyclopedia of marketing*. American Cancer Society, Georgia. <https://onlinelibrary.wiley.com/doi/abs/https://doi.org/10.1002/9781444316568.wiem01015>
- Kumar A, Garg G (2019) Systematic literature review on context-based sentiment analysis in social multimedia. *Multimed Tools Appl* 79:15349–15380
- Kumar R, Garg S (2020) Aspect-based sentiment analysis using deep learning convolutional neural network. *Information and communication technology for sustainable development*. Springer, Singapore, pp 43–52
- Kumar A, Jaiswal A (2020) Systematic literature review of sentiment analysis on Twitter using soft computing techniques. *Concurr Comput Pract Exp* 32(1):e5107
- Kumar NS, Malarvizhi N (2020) Bi-directional LSTM–CNN combined method for sentiment analysis in part of speech tagging (PoS). *Int J Speech Technol* 23:373–380
- Kumar V, Reinartz W (2016) Creating enduring customer value. *J Mark* 80(6):36–68. <https://doi.org/10.1509/jm.15.0414>
- Kumar A, Sebastian TM (2012) Sentiment analysis: a perspective on its past, present and future. *Int J Intell Syst Appl* 4(10):1–14. <https://doi.org/10.5815/ijisa.2012.10.01>
- Kumar A, Sharan A (2020) Deep learning-based frameworks for aspect-based sentiment analysis. *Deep learning-based approaches for sentiment analysis*. Springer, Singapore, pp 139–158
- Kumar A, Sharma A (2017) Systematic literature review on opinion mining of big data for government intelligence. *Webology* 14(2):6–47 (**Scopus**)
- Kumar R, Pannu HS, Malhi AK (2020) Aspect-based sentiment analysis using deep networks and stochastic optimization. *Neural Comput Appl* 32(8):3221–3235
- Kumar A, Srinivasan K, Cheng WH, Zomaya AY (2020) Hybrid context enriched deep learning model for fine-grained sentiment analysis in textual and visual semiotic modality social data. *Inf Process Manag* 57(1):102141

- Ładyżyński P, Żbikowski K, Gawrysiak P (2019) Direct marketing campaigns in retail banking with the use of deep learning and random forests. *Expert Syst Appl* 134:28–35. <https://doi.org/10.1016/j.eswa.2019.05.020>
- Lai Y, Zhang L, Han D, Zhou R, Wang G (2020) Fine-grained emotion classification of Chinese micro-blogs based on graph convolution networks. *World Wide Web* 23(5):2771–2787
- Li G, Liu F (2014) Sentiment analysis based on clustering: a framework in improving accuracy and recognizing neutral opinions. *Appl Intell* 40(3):441–452. <https://doi.org/10.1007/s10489-013-0463-3>
- Li F, Huang M, Yang Y, Zhu X (2011) Learning to identify review spam. In: *Proceedings of the twenty-second international joint conference on artificial intelligence-volume Vol 3*. pp 2488–2493
- Li W, Zhu L, Shi Y, Guo K, Zheng Y (2020) User reviews: Sentiment analysis using lexicon integrated two-channel CNN-LSTM family models. *Appl Soft Comput* 94(106435):1–11
- Li L, Goh TT, Jin D (2020) How textual quality of online reviews affect classification performance: a case of deep learning sentiment analysis. *Neural Comput Appl* 32(9):4387–4415
- Li D, Rzepka R, Ptaszynski M, Araki K (2020) HEMOS: a novel deep learning-based fine-grained humor detecting method for sentiment analysis of social media. *Inf Process Manag* 57(6):102290
- Lim WL, Ho CC, Ting CY (2020) Tweet sentiment analysis using deep learning with nearby locations as features. *Computational science and technology*. Springer, Singapore, pp 291–299
- Lin Y, Li J, Yang L, Xu K, Lin H (2020) Sentiment analysis with comparison enhanced deep neural network. *IEEE Access* 8:78378–78384
- Ling M, Chen Q, Sun Q, Jia Y (2020) Hybrid neural network for sina weibo sentiment analysis. *IEEE Trans Comput Soc Syst* 7(4):983–990
- Liu B (2020) Text sentiment analysis based on CBOW model and deep learning in big data environment. *J Ambient Intell Humaniz Comput* 11(2):451–458
- Liu Q, Mukaidani H (2020) Effective-target representation via LSTM with attention for aspect-level sentiment analysis. In: *2020 international conference on artificial intelligence in information and communication (ICAIIIC)*. IEEE, pp 336–340
- Liu N, Shen B (2020) Aspect-based sentiment analysis with gated alternate neural network. *Knowl-Based Syst* 188(105010):1–14
- Liu N, Shen B (2020) ReMemNN: a novel memory neural network for powerful interaction in aspect-based sentiment analysis. *Neurocomputing* 395:66–77
- Lou Y, Zhang Y, Li F, Qian T, Ji D (2020) Emoji-based sentiment analysis using attention networks. *ACM Trans Asian Low-Resour Lang Inf Process (TALLIP)* 19(5):1–13
- Lu Q, Zhu Z, Zhang D, Wu W, Guo Q (2020) Interactive rule attention network for aspect-level sentiment analysis. *IEEE Access* 8:52505–52516
- Lu G, Zhao X, Yin J, Yang W, Li B (2020) Multi-task learning using variational auto-encoder for sentiment classification. *Pattern Recogn Lett* 132:115–122
- Luo J, Huang S, Wang R (2020) A fine-grained sentiment analysis of online guest reviews of economy hotels in China. *J Hosp Mark Manag* 1–25
- Ma Y, Peng H, Cambria E (2018) Targeted aspect-based sentiment analysis via embedding common-sense knowledge into an attentive LSTM. In: *Proceedings of the AAAI conference on artificial intelligence vol 32*. pp 5876–5883
- Madhala P, Jussila J, Aramo-Immonen H, Suominen A (2018) Systematic literature review on customer emotions in social media. In: *ECSM 2018 5th European conference on social media*. Academic Conferences and publishing limited, South Oxfordshire, pp 154–162
- Maglogiannis IG (ed) (2007) *Emerging artificial intelligence applications in computer engineering: real word ai systems with applications in ehealth, hci, information retrieval and pervasive technologies*, vol 160. Ios Press, Amsterdam
- Mahmood Z, Safder I, Nawab RMA, Bukhari F, Nawaz R, Alfakeeh AS, Hassan SU (2020) Deep sentiments in Roman Urdu text using recurrent convolutional neural network model. *Inf Process Manag* 57(4):102233
- Majumder N, Poria S, Peng H, Chhaya N, Cambria E, Gelbukh A (2019) Sentiment and sarcasm classification with multitask learning. *IEEE Intell Syst* 34(3):38–43
- Meškelė D, Frasincar F (2020) ALDONAr: a hybrid solution for sentence-level aspect-based sentiment analysis using a lexicalized domain ontology and a regularized neural attention model. *Inf Process Manag* 57(3):102211
- Minaee S, Azimi E, Abdolrashidi A (2019) Deep-sentiment: sentiment analysis using ensemble of cnn and bi-lstm models. <http://arxiv.org/abs/arXiv:1904.04206>
- Minaee S, Kalchbrenner N, Cambria E, Nikzad N, Chenaghlu M, Gao J (2020) Deep learning based text classification: a comprehensive review 1(1):1–43. <http://arxiv.org/abs/arXiv:2004.03705>

- Mite-Baidal K, Delgado-Vera C, Solís-Avilés E, Espinoza AH, Ortiz-Zambrano J, Varela-Tapia E (2018) Sentiment analysis in education domain: a systematic literature review. *Commun Comput Inf Sci* 883:285–297. https://doi.org/10.1007/978-3-030-00940-3_21 (Scopus)
- Naseem U, Razzak I, Musial K, Imran M (2020) Transformer based deep intelligent contextual embedding for twitter sentiment analysis. *Future Gener Comput Syst* 113:58–69
- Nguyen TH, Shirai K (2015) Topic modeling based sentiment analysis on social media for stock market prediction. In: Proceedings of the 53rd annual meeting of the association for computational linguistics and the 7th international joint conference on natural language processing vol 1. pp 1354–1364. <https://doi.org/https://doi.org/10.3115/v1/P15-1131>
- Nigam K, Ghani R (2000) Analyzing the effectiveness and applicability of co-training. In: Proceedings of the ninth international conference on information and knowledge management—CIKM '00. pp 86–93. <https://doi.org/https://doi.org/10.1145/354756.354805>
- Nurdiani I, Börstler J, Fricker SA (2016) The impacts of agile and lean practices on project constraints: a tertiary study. *J Syst Softw* 119:162–183
- Ombabi AH, Ouarda W, Alimi AM (2020) Deep learning CNN–LSTM framework for Arabic sentiment analysis using textual information shared in social networks. *Soc Netw Anal Min* 10(1):1–13
- Onan A (2020) Mining opinions from instructor evaluation reviews: a deep learning approach. *Comput Appl Eng Edu* 28(1):117–138
- Onan A (2020a) Sentiment analysis on massive open online course evaluations: A text mining and deep learning approach. *Comput Appl Eng Edu* 1–18
- Onan A (2020b) Sentiment analysis on product reviews based on weighted word embeddings and deep neural networks. *Concurr Comput Pract Exp* e5909
- Ott M, Choi Y, Cardie C, Hancock JT (2011) Finding deceptive opinion spam by any stretch of the imagination. In: Proceedings of the 49th annual meeting of the association for computational linguistics: human language technologies vol 1. pp 309–319
- Pan Y, Liang M (2020) Chinese text sentiment analysis based on BI-GRU and self-attention. In: 2020 IEEE 4th information technology, networking, electronic and automation control conference (ITNEC) vol. 1. IEEE, pp 1983–1988
- Parimala M, Swarna Priya RM, Praveen Kumar Reddy M, Lal Chowdhary C, Kumar Poluru R, Khan S (2020) Spatiotemporal-based sentiment analysis on tweets for risk assessment of event using deep learning approach. *Softw Pract Exp* 1–21
- Park HJ, Song M, Shin KS (2020) Deep learning models and datasets for aspect term sentiment classification: implementing holistic recurrent attention on target-dependent memories. *Knowl-Based Syst* 187(104825):1–15
- Patel P, Patel D, Naik C (2020) Sentiment analysis on movie review using deep learning RNN method. *Intelligent data engineering and analytics*. Springer, Singapore, pp 155–163
- Pavlinek M, Podgorelec V (2017) Text classification method based on self-training and LDA topic models. *Expert Syst Appl* 80:83–93. <https://doi.org/10.1016/j.eswa.2017.03.020>
- Payne A, Frow P (2005) A strategic framework for customer relationship management. *J Mark* 69(4):167–176. <https://doi.org/10.1509/jmkg.2005.69.4.167>
- Peng M, Zhang Q, Jiang Y, Huang X (2018) Cross-domain sentiment classification with target domain specific information. In: Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics vol 1. pp 2505–2513. <https://doi.org/https://doi.org/10.18653/v1/P18-1233>
- 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 AAAI. pp 8600–8607
- Pergola G, Gui L, He Y (2019) TDAM: a topic-dependent attention model for sentiment analysis. *Inf Process Manag* 56(6):102084
- Petersen K, Feldt R, Mujtaba S, Mattsson M (2008) Systematic mapping studies in software engineering. In: 12th international conference on evaluation and assessment in software engineering (EASE) 12. pp 1–10
- Phillips-Wren G, Hoskisson A (2015) An analytical journey towards big data. *J Decis Syst* 24(1):87–102. <https://doi.org/10.1080/12460125.2015.994333>
- Poria S, Chaturvedi I, Cambria E, Bisio F (2016) Sentic LDA: improving on LDA with semantic similarity for aspect-based sentiment analysis. In: 2016 international joint conference on neural networks (IJCNN). IEEE, pp 4465–4473
- Poria S, Majumder N, Hazarika D, Cambria E, Gelbukh A, Hussain A (2018) Multimodal sentiment analysis: addressing key issues and setting up the baselines. *IEEE Intell Syst* 33(6):17–25
- Poria S, Hazarika D, Majumder N, Mihalcea R (2020) Beneath the tip of the iceberg: current challenges and new directions in sentiment analysis research. <http://arxiv.org/abs/arXiv:2005.00357>

- Portugal I, Alencar P, Cowan D (2018) The use of machine learning algorithms in recommender systems: a systematic review. *Expert Syst Appl* 97:205–227. <https://doi.org/10.1016/j.eswa.2017.12.020>
- Pozzi FA, Fersini E, Messina E, Liu B (2017) Challenges of sentiment analysis in social networks: an overview. In: Pozzi FA, Fersini E, Messina E, Liu B (eds) *Sentiment analysis in social networks*. Morgan Kaufmann, Burlington, pp 1–11
- Pröllochs N, Feuerriegel S, Lutz B, Neumann D (2020) Negation scope detection for sentiment analysis: a reinforcement learning framework for replicating human interpretations. *Inf Sci* 536:205–221
- Qazi A, Raj RG, Hardaker G, Standing C (2017) A systematic literature review on opinion types and sentiment analysis techniques: tasks and challenges. *Internet Res* 27(3):608–630. <https://doi.org/10.1108/IntR-04-2016-0086> (Scopus)
- Qiu G, Liu B, Bu J, Chen C (2009) Expanding domain sentiment lexicon through double propagation. In: *IJCAI* vol 9. pp 1199–1204
- Raatikainen M, Tiihonen J, Männistö T (2019) Software product lines and variability modeling: a tertiary study. *J Syst Softw* 149:485–510. <https://doi.org/10.1016/j.jss.2018.12.027>
- Rababah K, Mohd H, Ibrahim H (2011) A unified definition of CRM towards the successful adoption and implementation. *Acad Res Int* 1(1):220–228
- Rambocas M, Pacheco BG (2018) Online sentiment analysis in marketing research: a review. *J Res Interact Mark* 12(2):146–163. <https://doi.org/10.1108/JRIM-05-2017-0030>
- Rao AVSR, Ranjana P (2020) Deep learning method to identify the demographic attribute to enhance effectiveness of sentiment analysis. *Innovations in computer science and engineering*. Springer, Singapore, pp 275–285
- Ravi K, Ravi V (2015) A survey on opinion mining and sentiment analysis: tasks, approaches and applications. *Knowl-Based Syst* 89:14–46. <https://doi.org/10.1016/j.knosys.2015.06.015>
- Ray P, Chakrabarti A (2020) A mixed approach of deep learning method and rule-based method to improve aspect level sentiment analysis. *Appl Comput Inform*. <https://doi.org/10.1016/j.aci.2019.02.002>
- Reddy YCAP, Viswanath P, Eswara Reddy B (2018) Semi-supervised learning: a brief review. *Int J Eng Technol* 7(18):81
- Reichheld FF, Schefter P (2000) E-loyalty: your secret weapon on the web. *Harv Bus Rev* 78(4):105–113
- Reinartz W, Krafft M, Hoyer WD (2004) The customer relationship management process: its measurement and impact on performance. *J Mark Res* 41(3):293–305. <https://doi.org/10.1509/jmkr.41.3.293.35991>
- Ren Z, Zeng G, Chen L, Zhang Q, Zhang C, Pan D (2020) A lexicon-enhanced attention network for aspect-level sentiment analysis. *IEEE Access* 8:93464–93471
- Ren F, Feng L, Xiao D, Cai M, Cheng S (2020) DNet: a lightweight and efficient model for aspect based sentiment analysis. *Expert Syst Appl* 151(113393):1–10
- Ren L, Xu B, Lin H, Liu X, Yang L (2020) Sarcasm detection with sentiment semantics enhanced multi-level memory network. *Neurocomputing* 401:320–326
- Rios N, de Mendonça Neto MG, Spínola RO (2018) A tertiary study on technical debt: types, management strategies, research trends, and base information for practitioners. *Inf Softw Technol* 102:117–145. <https://doi.org/10.1016/j.infsof.2018.05.010>
- Rodrigues Chagas BN, Nogueira Viana JA, Reinhold O, Lobato F, Jacob AFL, Alt R (2018) Current applications of machine learning techniques in CRM: a literature review and practical implications. *IEEE/WIC/ACM Int Conf Web Intell (WI)* 2018:452–458. <https://doi.org/10.1109/WI.2018.00-53>
- Rojas-Barahona LM (2016) Deep learning for sentiment analysis. *Lang Linguist Compass* 10(12):701–719
- Rout JK, Dalmia A, Choo K-KR, Bakshi S, Jena SK (2017) Revisiting semi-supervised learning for online deceptive review detection. *IEEE Access* 5:1319–1327. <https://doi.org/10.1109/ACCESS.2017.2655032>
- Rygielski C, Wang J-C, Yen DC (2002) Data mining techniques for customer relationship management. *Technol Soc* 24(4):483–502. [https://doi.org/10.1016/S0160-791X\(02\)00038-6](https://doi.org/10.1016/S0160-791X(02)00038-6)
- Sabbeh SF (2018) Machine-learning techniques for customer retention: A comparative study. *Int J Adv Comput Sci Appl* 9(2):273–281
- Sadr H, Pedram MM, Teshnehlab M (2020) Multi-view deep network: a deep model based on learning features from heterogeneous neural networks for sentiment analysis. *IEEE Access* 8:86984–86997
- Salah Z, Al-Ghuwairi A-RF, Baarah A, Aloqaily A, Qadoumi B, Alhayek M, Alhijawi B (2019) A systematic review on opinion mining and sentiment analysis in social media. *Int J Bus Inf Syst* 31(4):530–554. <https://doi.org/10.1504/IJBIS.2019.101585> (Scopus)
- Salur MU, Aydin I (2020) A novel hybrid deep learning model for sentiment classification. *IEEE Access* 8:58080–58093
- Sangeetha K, Prabha D (2020) Sentiment analysis of student feedback using multi-head attention fusion model of word and context embedding for LSTM. *J Ambient Intell Humaniz Comput*. <https://doi.org/10.1007/s12652-020-01791-9>

- Sankar H, Subramaniaswamy V, Vijayakumar V, Arun Kumar S, Logesh R, Umamakeswari AJSP (2020) Intelligent sentiment analysis approach using edge computing-based deep learning technique. *Softw Pract Exp* 50(5):645–657
- Sawant SS, Prabukumar M (2018) A review on graph-based semi-supervised learning methods for hyperspectral image classification. *Egypt J Remote Sens Space Sci* 23(2):243–248. <https://doi.org/10.1016/j.ejrs.2018.11.001>
- Schouten K, Frasincar F (2016) Survey on aspect-level sentiment analysis. *IEEE Trans Knowl Data Eng* 28(3):813–830. <https://doi.org/10.1109/TKDE.2015.2485209>
- Seo S, Kim C, Kim H, Mo K, Kang P (2020) Comparative study of deep learning-based sentiment classification. *IEEE Access* 8:6861–6875
- Shakeel MH, Karim A (2020) Adapting deep learning for sentiment classification of code-switched informal short text. In: *Proceedings of the 35th annual ACM symposium on applied computing*. pp 903–906
- Sharma SS, Dutta G (2018) Polarity determination of movie reviews: a systematic literature review. *Int J of Innov Knowl Concepts* 6:12
- Shayaa S, Jaafar NI, Bahri S, Sulaiman A, Seuk Wai P, Wai Chung Y, Piprani AZ, Al-Garadi MA (2018) Sentiment analysis of big data: Methods, applications, and open challenges. *IEEE Access* 6:37807–37827. <https://doi.org/10.1109/ACCESS.2018.2851311> (Scopus)
- Shirani-Mehr H (2014) Applications of deep learning to sentiment analysis of movie reviews. Tech Report 1–8
- Shuang K, Yang Q, Loo J, Li R, Gu M (2020) Feature distillation network for aspect-based sentiment analysis. *Inf Fusion* 61:13–23
- Silva NFFD, Coletta LFS, Hruschka ER (2016) A survey and comparative study of tweet sentiment analysis via semi-supervised learning. *ACM Comput Surv* 49(1):1–26. <https://doi.org/10.1145/2932708>
- Singh PK, Sharma S, Paul S (2020) Identifying hidden sentiment in text using deep neural network. In *2nd international conference on data, engineering and applications (IDEA)*. IEEE, pp 1–5
- Socher R, Perelygin A, Wu J, Chuang J, Manning CD, Ng AY, 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
- Studiawan H, Sohel F, Payne C (2020) Sentiment analysis in a forensic timeline with deep learning. *IEEE Access* 8:60664–60675
- Su YJ, Hu WC, Jiang JH, Su RY (2020) A novel LMAEB-CNN model for Chinese microblog sentiment analysis. *J Supercomput* 76:9127–9141
- Sun X, He J (2020) A novel approach to generate a large scale of supervised data for short text sentiment analysis. *Multimed Tools Appl* 79(9):5439–5459
- Taboada M, Brooke J, Tofiloski M, Voll K, Stede M (2011) Lexicon-based methods for sentiment analysis. *Comput Linguist* 37(2):267–307
- Tang, D., Wei, F., Yang, N., Zhou, M., Liu, T., & Qin, B. (2014, June). Learning sentiment-specific word embedding for twitter sentiment classification. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics vol 1*. pp 1555–1565
- Tao J, Fang X (2020) Toward multi-label sentiment analysis: a transfer learning based approach. *J Big Data* 7(1):1–26
- Thet TT, Na J-C, Khoo CSG (2010) Aspect-based sentiment analysis of movie reviews on discussion boards. *J Inf Sci* 36(6):823–848. <https://doi.org/10.1177/0165551510388123>
- Tran TU, Hoang HTT, Huynh HX (2020) Bidirectional independently long short-term memory and conditional random field integrated model for aspect extraction in sentiment analysis. *Frontiers in intelligent computing: theory and applications*. Springer, Singapore, pp 131–140
- Tsai C, Hu Y, Hung C, Hsu Y (2013) A comparative study of hybrid machine learning techniques for customer lifetime value prediction. *Kybernetes* 42(3):357–370. <https://doi.org/10.1108/03684921311323626>
- Ullah MA, Marium SM, Begum SA, Dipa NS (2020) An algorithm and method for sentiment analysis using the text and emoticon. *ICT Express* 6(4):357–360
- Usama M, Ahmad B, Song E, Hossain MS, Alrashoud M, Muhammad G (2020) Attention-based sentiment analysis using convolutional and recurrent neural network. *Future Gener Comput Syst* 113:571–578
- Valdivia A, Luzón MV, Herrera F (2017) Sentiment analysis in tripadvisor. *IEEE Intell Syst* 32(4):72–77
- Valdivia A, Martínez-Cámara E, Chaturvedi I, Luzón MV, Cambria E, Ong YS, Herrera F (2020) What do people think about this monument? understanding negative reviews via deep learning, clustering and descriptive rules. *J Ambient Intell Humaniz Comput* 11(1):39–52
- Vechtomova O (2017) Disambiguating context-dependent polarity of words: an information retrieval approach. *Inf Process Manag* 53(5):1062–1079
- Venkatakrishnan S, Kaushik A, Verma JK (2020) Sentiment analysis on google play store data using deep learning. *Applications of machine learning*. Springer, Singapore, pp 15–30

- Verhoef PC, Venkatesan R, McAlister L, Malthouse EC, Krafft M, Ganesan S (2010) CRM in data-rich multichannel retailing environments: a review and future research directions. *J Interact Mark* 24(2):121–137. <https://doi.org/10.1016/j.intmar.2010.02.009>
- Verner JM, Brereton OP, Kitchenham BA, Turner M, Niazi M (2014) Risks and risk mitigation in global software development: a tertiary study. *Inf Softw Technol* 56(1):54–78
- Vinodhini G, Chandrasekaran RM (2012) Sentiment analysis and opinion mining: a survey. *Int J* 2(6):282–292
- Vosoughi S, Roy D, Aral S (2018) The spread of true and false news online. *Science* 359(6380):1146–1151
- Vyas V, Uma V (2019) Approaches to sentiment analysis on product reviews. *Sentiment analysis and knowledge discovery in contemporary business*. IGI Global, Pennsylvania, pp 15–30
- Wadawadagi R, Pagi V (2020) Sentiment analysis with deep neural networks: comparative study and performance assessment. *Artif Intell Rev* 53:6155–6195
- Wang G, Sun J, Ma J, Xu K, Gu J (2014) Sentiment classification: the contribution of ensemble learning. *Decis Support Syst* 57:77–93. <https://doi.org/10.1016/j.dss.2013.08.002>
- Wang Y, Huang M, Zhu X, Zhao L (2016) Attention-based LSTM for aspect-level sentiment classification. In: *Proceedings of the 2016 conference on empirical methods in natural language processing*. pp 606–615
- Wang S, Zhu Y, Gao W, Cao M, Li M (2020) Emotion-semantic-enhanced bidirectional LSTM with multi-head attention mechanism for microblog sentiment analysis. *Information* 11(5):280
- Wehrmann J, Becker W, Cagnini HE, Barros RC (2017) A character-based convolutional neural network for language-agnostic Twitter sentiment analysis. In: *2017 International joint conference on neural networks (IJCNN)*. IEEE, pp 2384–2391
- Wilcox PA, Gurău C (2003) Business modelling with UML: the implementation of CRM systems for online retailing. *J Retail Consum Serv* 10(3):181–191. [https://doi.org/10.1016/S0969-6989\(03\)00004-3](https://doi.org/10.1016/S0969-6989(03)00004-3)
- Winer RS (2001) A framework for customer relationship management. *Calif Manag Rev* 43(4):89–105. <https://doi.org/10.2307/41166102>
- Wu C, Wu F, Wu S, Yuan Z, Liu J, Huang Y (2019) Semi-supervised dimensional sentiment analysis with variational autoencoder. *Knowl-Based Syst* 165:30–39
- Xi D, Zhuang F, Zhou G, Cheng X, Lin F, He Q (2020) Domain adaptation with category attention network for deep sentiment analysis. In: *Proceedings of the web conference 2020*. pp 3133–3139
- Xia Y, Cambria E, Hussain A, Zhao H (2015) Word polarity disambiguation using bayesian model and opinion-level features. *Cognit Comput* 7(3):369–380
- Xu W, Tan Y (2019) Semi-supervised target-oriented sentiment classification. *Neurocomputing* 337:120–128
- Xu G, Meng Y, Qiu X, Yu Z, Wu X (2019) Sentiment analysis of comment texts based on BiLSTM. *IEEE Access* 7:51522–51532
- Yadav A, Vishwakarma DK (2020) Sentiment analysis using deep learning architectures: a review. *Artif Intell Rev* 53(6):4335–4385
- Yadav A, Vishwakarma DK (2020) A deep learning architecture of RA-DLNet for visual sentiment analysis. *Multimed Syst* 26:431–451
- Yang L, Li Y, Wang J, Sherratt RS (2020) Sentiment analysis for E-commerce product reviews in chinese based on sentiment lexicon and deep learning. *IEEE Access* 8:23522–23530
- Yao F, Wang Y (2020) Domain-specific sentiment analysis for tweets during hurricanes (DSSA-H): a domain-adversarial neural-network-based approach. *Comput Environ Urban Syst* 83(101522):1–14
- Yarowsky D (1995) Unsupervised word sense disambiguation rivaling supervised methods. In: *Proceedings of the 33rd annual meeting on association for computational linguistics*. pp189–196. <https://doi.org/https://doi.org/10.3115/981658.981684>
- Yildirim S (2020) Comparing deep neural networks to traditional models for sentiment analysis in Turkish language. *Deep learning-based approaches for sentiment analysis*. Springer, Singapore, pp 311–319
- Zerbino P, Aloini D, Dulmin R, Mininno V (2018) Big Data-enabled customer relationship management: a holistic approach. *Inf Process Manag* 54(5):818–846. <https://doi.org/10.1016/j.ipm.2017.10.005>
- Zhang L, Wang S, Liu B (2018) Deep learning for sentiment analysis: a survey. *Wiley Interdiscip Rev Data Min Knowl Discov* 8(4):e1253
- Zhang B, Li X, Xu X, Leung KC, Chen Z, Ye Y (2020) Knowledge guided capsule attention network for aspect-based sentiment analysis. *IEEE/ACM Trans Audio Speech Lang Process* 28:2538–2551
- Zhang S, Xu X, Pang Y, Han J (2020) Multi-layer attention based CNN for target-dependent sentiment classification. *Neural Process Lett* 51(3):2089–2103
- Zhao P, Hou L, Wu O (2020) Modeling sentiment dependencies with graph convolutional networks for aspect-level sentiment classification. *Knowl-Based Syst* 193(105443):1–10

- Zhou J, Huang JX, Hu QV, He L (2020) Is position important? deep multi-task learning for aspect-based sentiment analysis. *Appl Intell* 50:3367–3378
- Zhu X, Ghahramani Z (2002) Learning from labeled and unlabeled data with label propagation. Technical report CMU-CALD-02-107, Carnegie Mellon University. 8
- Zhu X, Yin S, Chen Z (2020) Attention based BiLSTM-MCNN for sentiment analysis. In: 2020 IEEE 5th international conference on cloud computing and big data analytics (ICCCBDA). IEEE, pp 170–174
- Zuo E, Zhao H, Chen B, Chen Q (2020) Context-specific heterogeneous graph convolutional network for implicit sentiment analysis. *IEEE Access* 8:37967–37975

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