

# Sentiment analysis using deep learning approaches: an overview

Olivier HABIMANA<sup>1</sup>, Yuhua LI<sup>1\*</sup>, Ruixuan LI<sup>1\*</sup>, Xiwu GU<sup>1</sup> & Ge YU<sup>2</sup>

<sup>1</sup>*School of Computer Science and Technology, Huazhong University of Science and Technology, Wuhan 430074, China;*

<sup>2</sup>*School of Computer Science and Engineering, Northeastern University, Shenyang 110819, China*

Received 18 August 2018/Revised 2 December 2018/Accepted 4 June 2019/Published online 26 December 2019

**Abstract** Nowadays, with the increasing number of Web 2.0 tools, users generate huge amounts of data in an enormous and dynamic way. In this regard, the sentiment analysis appeared to be an important tool that allows the automation of getting insight from the user-generated data. Recently, deep learning approaches have been proposed for different sentiment analysis tasks and have achieved state-of-the-art results. Therefore, in order to help researchers to depict quickly the current progress as well as current issues to be addressed, in this paper, we review deep learning approaches that have been applied to various sentiment analysis tasks and their trends of development. This study also provides the performance analysis of different deep learning models on a particular dataset at the end of each sentiment analysis task. Toward the end, the review highlights current issues and hypothesized solutions to be taken into account in future work. Moreover, based on knowledge learned from previous studies, the future work subsection shows the suggestions that can be incorporated into new deep learning models to yield better performance. Suggestions include the use of bidirectional encoder representations from transformers (BERT), sentiment-specific word embedding models, cognition-based attention models, common sense knowledge, reinforcement learning, and generative adversarial networks.

**Keywords** sentiment analysis, opinion mining, deep learning, neural network, natural language processing (NLP), social network

**Citation** Habimana O, Li Y H, Li R X, et al. Sentiment analysis using deep learning approaches: an overview. *Sci China Inf Sci*, 2020, 63(1): 111102, <https://doi.org/10.1007/S11432-018-9941-6>

## 1 Introduction

Recently, there is a remarkable increase in the number of Web 2.0 tools like online social media and e-commerce websites where users freely express their ideas and thoughts. Owing to this increase, huge amounts of data are generated. Therefore, sentiment analysis was introduced as a tool for automatic extraction of insight and useful information from the user-generated data [1]. Sentiment analysis is one of the natural language processing (NLP) tasks. It has attracted a large number of researchers and industry communities because of its usefulness and challenges [2, 3].

Sentiment analysis is a field of study whose main objective is to identify and examine the components of a person's opinion. According to the definition by Liu [2, 4], an opinion normally consists of an entity, aspects of an entity, and the sentiment of aspect that represents its polarity. Furthermore, it includes the author of the opinion and the time when the opinion is expressed. The sentiments of an aspect are classified into different categories depending on the purpose of sentiment classification. For example, Yang et al. [5] classified tweets into positive, negative and neutral classes. Similarly, Kalchbrenner et

\* Corresponding author (email: [icluyuhua@hust.edu.cn](mailto:icluyuhua@hust.edu.cn), [rxli@hust.edu.cn](mailto:rxli@hust.edu.cn))

al. [6] classified the movie reviews into negative, somewhat negative, neutral, somewhat positive, and positive classes.

In past years, several approaches for sentiment analysis have been proposed and pioneered by Stone et al. [7] who developed a computer system for content analysis using lexicon. Later, Pang et al. [8] applied machine learning to find the sentiment expressed in text data. These traditional approaches have achieved good results but the feature engineering they rely on is a tedious task. Later, researchers realized that finding the sentiment for today's user-generated data requires deep understanding and efficient methods are needed to cope with it. Therefore, deep learning (DL) approaches evolved as efficient methods due to their capability of learning the text without manual feature engineering. DL approaches have been proved to outperform the traditional methods in sentiment analysis. Thus, we review the recent DL models that have been proposed for various tasks of sentiment analysis.

In contrast to existing surveys of sentiment analysis using DL methods, we discuss the most recent DL approaches and their variants that have been applied to different sentiment analysis tasks. The recent survey [9] has reviewed DL approaches for sentiment analysis but our work is conducted from a different perspective. In addition, it did not address trending DL methods like deep reinforcement learning (DRL) and generative adversarial networks (GANs). Rojas-Barahona et al. [10] emphasized on the technical overview of DL for sentiment analysis but they did not discuss most variants of DL approaches and recent studies. Therefore, we are sure that this review will give readers a complete idea of recent trends, with recent studies and new methods in sentiment analysis using DL approaches such as traditional attention mechanism, cognitive attention based models, DRL models, and GANs. Furthermore, different from the above-cited reviews, readers will get an introduction of sentiment-specific word embedding models and different datasets used in sentiment analysis. Moreover, to help readers quickly get the top performing models on a particular task and dataset, we provide the performance analysis of different DL models for different tasks of sentiment analysis. At the end of this paper, readers will know the recent research issues that need to be addressed and the future directions.

The rest of the review is structured as follows. Section 2 provides the background of sentiment analysis. In Section 3, we discuss word embedding models and offer a quick introduction of different DL approaches that are commonly applied in sentiment analysis. Section 4 introduces various real-world datasets that are commonly used in sentiment analysis. In Section 5, we discuss the applications and performance of DL approaches on various tasks of sentiment analysis. The current issues that are worthy to be addressed and future directions are highlighted in Section 6. Finally, Section 7 concludes the review with final remarks.

## 2 Background of sentiment analysis

### 2.1 Application of sentiment analysis

The need for sentiment analysis ranges from individuals to large organizations and governments. An individual customer first checks the ratings of a product and opinions from other customers before making his/her purchasing decision, and business organizations use sentiment analysis tools to understand their customers' feeling. Several efforts [11–14] have been made in applying sentiment analysis to customer reviews. Therefore, analyzing customer reviews have been proved to boost the relevant market and increase the confidence of customers [15]. In addition, the governments analyze the feelings of the public about trending topics like elections and their policies. A recent case is the prediction of the 2016 USA presidential election [16]. Furthermore, sentiment analysis can be used to enhance the capability of recommendation systems where users' interests can be identified [17,18].

### 2.2 Tasks of sentiment analysis

In satisfying the need of different individuals and organizations, sentiment analysis consists of five main tasks [2,4], which are based on the five components of the opinion identified in Section 1. The first

task deals with entity extraction and categorization. This task is perceived as named entity recognition (NER) that was first introduced in 2005 [19]. Also, in 2004, Hu et al. [20] introduced the second task, which involves aspect extraction and categorization. To address sentiment or opinion classification, the third task was popularized in 2002 [8, 21]. In 2004, Kim et al. [22] started the fourth task that handles opinion holder extraction and categorization. The fifth task, also considered as NER, was proposed to deal with time extraction and standardization. In the last task, all five components are extracted at the same time. In addition to the aforesaid tasks, Pozzi et al. [1] consolidated other tasks including emotion detection suggested in 2005 [23], opinion spam detection instigated in 2008 [24], multi-lingual sentiment analysis initiated in 2009 [25], multi-modal sentiment analysis introduced in 2011 [26] and opinion summarization [20]. Furthermore, there are other tasks like sentiment dynamic tracking [27] popularized in 2012 and sentiment collocation [28] that involves the extraction of the targets and related opinion terms based on their correlation.

### 2.3 Levels of sentiment analysis

Conducting sentiment analysis is more than classifying a document or a sentence into positive or negative classes. Indeed, finding the sentiment discussed in every aspect or feature of the entity is of prime importance. Therefore, depending upon the granularity required, the sentiment analysis task is performed at the document, sentence and aspect level. You can refer to [1, 2, 4] for detailed explanations about levels of sentiment analysis.

In late 2002, the concept of document level sentiment analysis was popularized by [8, 21], which mainly focuses on finding the polarity of the whole opinionated document with respect to a single entity (e.g., classifying the whole review document which talks about a given T-shirt). Without hesitation, we express our gratitude to this task for popularizing sentiment analysis. However, document level sentiment analysis later proved to be limited for providing enough information, as it does not consider different sentences and aspects that a document may contain. Thus, in 2004, sentence level sentiment analysis appeared to alleviate the problem by assigning the polarity to each opinionated sentence in the document [22]. For example, the sentence “This is a very good movie I saw!!” is classified as expressing positive sentiment. However, sentence level sentiment analysis suffers the same problem as the document level sentiment analysis for not providing enough information on what the customer really likes or dislikes, because a sentence may contain multiple entities with different aspects.

Consequently, to address the issue of dealing with multiple entities with different aspects, Hu et al. [20] in 2004 started a new milestone of feature level sentiment analysis by introducing feature-based opinion mining and summarization technique, which takes into account different product’s aspects and their sentiment polarity. Nowadays, sentiment analysis at this level is commonly known as aspect based sentiment analysis (ABSA) [4]. Specifically, it analyzes different features of an entity and finds exactly what someone likes or dislikes about that entity or aspect of the entity in a discussion. Considering the following example, “The food is good but the service is bad”, we can see clearly that the sentence is positive with respect to the aspect “food” but is negative with respect to the aspect “service”. So, ABSA is one of the most challenging yet highly needed task of sentiment analysis [2]. Moreover, today’s real-life applications of sentiment analysis are based on this level [4]. Thus, ABSA embodies many subtasks including aspect term extraction (ATE) [29], aspect-term sentiment analysis (ATSA) and aspect-category sentiment analysis (ACSA) [30].

### 2.4 Traditional approaches for sentiment analysis

The industrial and academic community have paid attention to sentiment analysis in order to assist in decision making. Consequently, researchers have proposed a large number of approaches in order to satisfy the need of sentiment analysis. The proposed traditional approaches have been proved to get good results by proper feature engineering. Thus, the commonly used features by these approaches in sentiment analysis are: part of speech (POS) tags [8], term position [8], opinion words and sentences [6, 31–33], negation [33], term presence and frequency [34], and syntactic dependency [35]. In quest of details about

these features, you can refer to [2, 8]. Therefore, this subsection is a brief overview of different traditional approaches for sentiment analysis with their limitations and their source references. The traditional approaches for sentiment analysis are classified into two categories: lexicon-based and machine learning approaches [2, 4, 36].

The lexicon-based approaches are types of the traditional approaches for sentiment analysis that use precompiled sentiment lexicons containing different words and their polarity to classify a given word into positive or negative sentiment class labels. The studies [2, 4, 36, 37] provide a detailed description of these approaches. Stone et al. [7] started the task of sentiment analysis using the lexicon method in 1966. Later, different lexicons were proposed such as WordNet, WordNet-Affect, SenticNet, MPQA, and SentiWordNet [31]. Following the popularity of lexicons, extensive research has been done for sentiment analysis based on lexicons [6, 31, 33, 38]. These approaches do not require the training dataset. However, the construction of the sentiment lexicon construction for today's user-generated unstructured data is a challenging task. Consequently, machine learning approaches help to alleviate the problem.

Machine learning approaches are other traditional methods for sentiment analysis that are based on the machine learning algorithms to classify the words into their corresponding sentiment labels. The main benefit of machine learning approaches is their ability of representation learning. Pang et al. [8] pioneered the use of these techniques for sentiment analysis. The surveys [2, 4, 36] give detailed explanations of these approaches. Machine learning algorithms require the training dataset which helps to automate the classifier and test dataset used for checking the operability of the classifier. Therefore, machine learning approaches are preferred for sentiment analysis due to their capacity for dealing with large amounts of data compared with lexicon based approaches [39]. However, in case there are no human annotated datasets, the majority of people choose to use lexicon based approaches [40]. Thus, extensive research has been done for sentiment analysis using machine learning approaches [41–43]. As far as the good results for sentiment analysis are concerned, the two traditional approaches for sentiment analysis can be combined in order to gain the advantages of each approach. Thus, the recent advents [44–46] are results for the combination of the two traditional approaches.

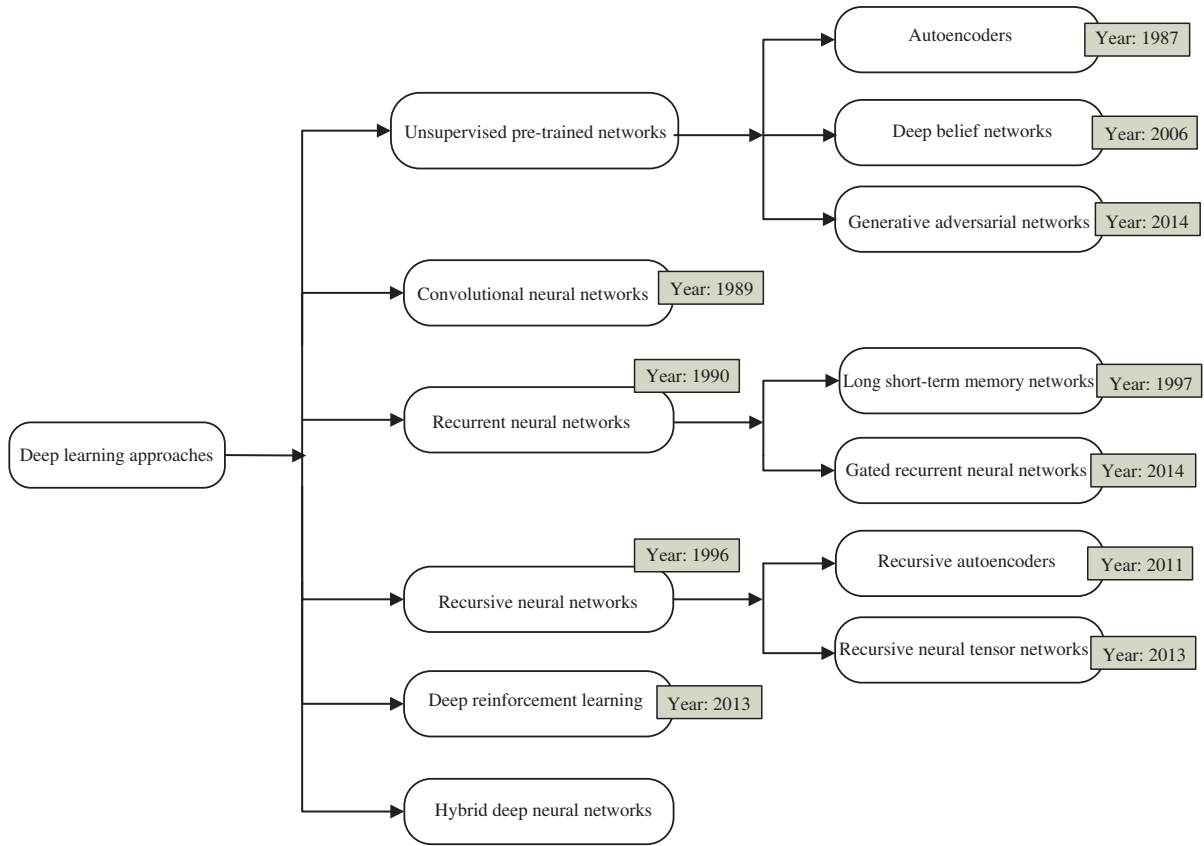
However, the traditional approaches for sentiment analysis are accused of being inefficient to cope with the new trend of data with dynamic nature of language, increase of high dimensional data, structural and cultural subtleties of short text like tweets. However, it is still a challenge to these traditional approaches to adjust a designed model for a specific task to a new task, especially lexicon based methods. Moreover, as mentioned in the introduction section, these approaches are based on the features engineering, which has been proved to be a tedious task with the present data [2]. In addition, multiclass classification leads the performance of traditional approaches to be poor, degraded and limited [2]. Thus, researchers mentioned the need of new approaches to solve these trending issues.

Consequently, to cope with the new trend of data, where efficient approaches for sentiment analysis are needed, researchers realized that the DL approaches give incredible results as affirmed by [5, 47, 48], and hence they are adopted for sentiment analysis.

### 3 Deep learning approaches

#### 3.1 Introduction

DL is an emerging branch of machine learning algorithms, which is inspired by artificial neural networks. It offers ways of learning the data representations in a supervised and unsupervised way with the help of the hierarchy of layers, which allow multiple processing [49, 50]. Foremost, the adoption of DL approaches in sentiment analysis has been driven by their ability of automatic feature learning, where they can learn automatically and discover discriminative and exploratory input representations from data themselves [8]. Moreover, their adoption has been motivated by the increase of the training data with multiclass classification and the success of word embeddings [51, 52]. Besides, the availability of powerful computing resources like the graphics processing unit (GPU) that allows efficient matrix manipulation [53] has also become the driving force to embrace DL approaches.



**Figure 1** (Color online) Deep learning approaches for sentiment analysis.

Recently, researchers have proposed a large number of DL approaches, and most of them have been applied to sentiment analysis. These approaches have been proved to be effective methods in sentiment analysis, which is evidenced by many studies that have been successfully done. They have solved complex issues like domain adaptation, being able deal with the context in which the word appears, and to model long-range dependencies which can change the polarity of a statement in a given sentence [54]. Consequently, this section first introduces word embedding approaches as the first data processing layer in DL methods and then proceed to various DL approaches with their trends of development. This study groups DL approaches into six categories: unsupervised pre-trained networks (UPNs), convolutional neural networks (CNNs), recurrent neural networks (RNNs), recursive neural networks (RvNNs), DRL, and hybrid neural networks. We bring to the readers' attention that in this review we mainly talk about two new DL approaches: GANs and DRLs, and we recommend the readers to survey [9] for the graphical and mathematical details of remaining DL approaches. Accordingly, Figure 1 shows the detailed classification of the above-mentioned approaches and their proposed years.

### 3.2 Word embeddings

Word embeddings are types of word representation that aim at representing words' meaning in the form of vectors, where words with similar meaning and context are represented by similar vectors. Word embeddings are considered as important ingredients in sentiment analysis as well as in other NLP tasks, and they serve as first data processing layer in DL approaches [54, 55]. Therefore, this subsection introduces different word embeddings that are commonly used in sentiment analysis.

Recent word embeddings follow the distribution hypothesis [56], where the words with the same context have similar meanings. Thus, the words with the same context or similar semantics create similar features and are classified in one class. Bengio et al. [57] initiated word embeddings by designing a language model, which learns distributed representation for each word and the likelihood function for word sequences at

the same time. Following the success of this model, extensive approaches have been suggested to improve the results and capture semantic and syntactic information. Collobert et al. [58] constructed a pre-trained word embedding model based on the DL model, which can learn the features necessary to a specific task when the prior information is not enough. The latter work has inspired and laid the foundation of many recent studies in the domain.

Consequently, different pre-trained word embeddings have been proposed and are available to the public. The popular pre-trained word embeddings used in the sentiment analysis are results of the combination of different models for word representations. Accordingly, Mikolov et al. [59] proposed word2vec<sup>1)</sup>, which is the result of the combination of the skip-gram and continuous bag-of-words (CBOW). The CBOW model predicts the center word from its surrounding context words whereas the skip-gram model predicts surrounding context words when given a center word. Another popular word embedding model GloVe<sup>2)</sup> means global vector has been developed by the NLP group at Stanford University [60]. GloVe combines global matrix decomposition and local context window. Similarly, Joulin et al. [61] proposed fasttext<sup>3)</sup> word embedding method where each word is represented by the character n-gram. Furthermore, these word embeddings can be refined using another word embedding. For example, Zou et al. [62] constructed a model that can be applied to both GloVe and word2vec for the purpose of capturing both semantic and sentimental information of the words. Moreover, recently NLP has received a breakthrough bidirectional encoder representations from transformers (BERT)<sup>4)</sup> language model that produces contextualized representation learning for different NLP tasks [63]. BERT makes use of transformer [64], which is based on a self-attention mechanism that explores the contextual relationship between words or sub-words that make input text.

However, the above word embeddings represent the word distributions without taking into account any specific task. Therefore, to address the problem, researchers have recently suggested word embedding models, which are tailored to sentiment analysis. Tang et al. [65] developed sentiment-specific word embedding (SSWE) that takes into account sentiment information whereas Zhou et al. [66] suggested a word embedding model that works in a cross-lingual setting to solve the problem of the semantic gap between English-Chinese for sentiment classification. Besides, Fu et al. [67] designed an SSWE that combines both local context and global sentiment information. Therefore, these characteristics of effectively understanding and representing the structure of words in the form of vectors lead the DL approaches to succeed in sentiment analysis.

Thus, the following subsections provide a brief introduction of DL approaches that are applied in sentiment analysis.

### 3.3 Unsupervised pre-trained networks

UPNs are types of deep neural networks that allow the unsupervised algorithm to pre-train the layers of the network with unlabeled data, and then this becomes the initial state for the final stage. And in the final stage, the network is fine-tuned with the supervised training [68]. Thus, the pre-training phase enables the network to converge quickly and learns the features from unlabeled data. Therefore, this capability of UPNs to learn features from unlabeled data through the reconstruction [69] has driven efforts to apply them in sentiment analysis. The commonly used UPNs in sentiment analysis are autoencoders (AEs), deep belief networks (DBNs), and GANs.

#### 3.3.1 Autoencoders

In 1987, Ballard [70] introduced autoencoders that allow the mapping of inputs to their outputs. Autoencoders are trained to reconstruct their inputs by minimizing the reconstruction error. Generally, they offer the following advantages: they can be used as standalone networks; they can serve as basic building

---

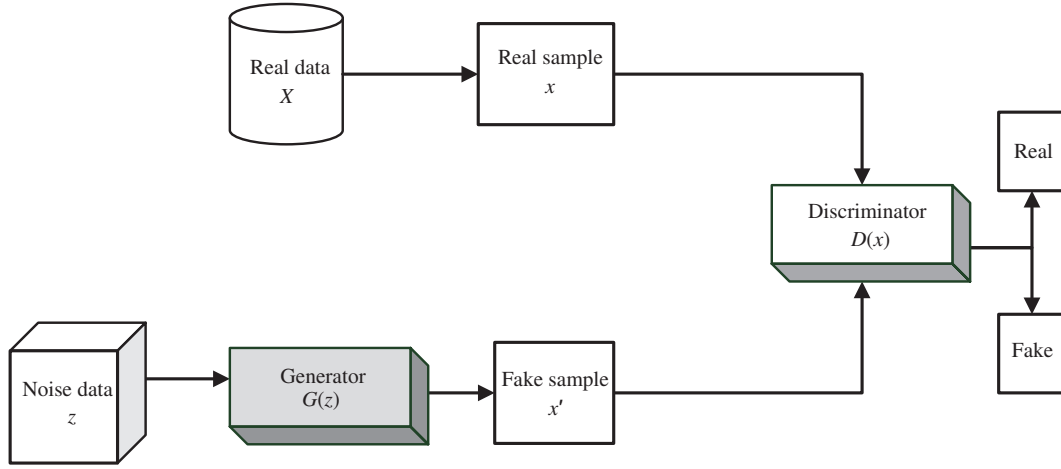
1) <https://github.com/tmikolov/word2vec>.

2) <https://nlp.stanford.edu/projects/glove/>.

3) <https://fasttext.cc/>.

4) <https://github.com/google-research/bert>.





**Figure 2** GAN framework with  $G$  and  $D$  during the training process.

blocks of other DL architectures [71]. In addition, autoencoders are considered as feature learning and dimensionality reduction techniques. In sentiment analysis, the commonly applied variants of autoencoders are denoising autoencoders (DAs) [72] introduced in 2008 and stacked denoising autoencoders (SDAs). The DAs have shown the capability to increase the robustness to noisy data and to reconstruct corrupted data. Particularly, in sentiment analysis, DAs were applied in domain adaptation by training a model on a labeled dataset and apply it on another unlabeled dataset. However, one layer of DA is not enough to obtain good representations because it could increase the reconstruction error. Therefore, a good solution is to stack different DAs (SDAs) because they increase the depth of a network, which is good for obtaining good representations. Thus, SDA helps to keep the number of training parameters as low as possible. Recently, some methods have been proposed in sentiment analysis, e.g., Rong et al. [73] proposed an autoencoder based method to address the issue of curse dimensionality in text document. Similarly, Zhou et al. [74] designed a cross-lingual sentiment analysis model, which eliminates the gap between English and Chinese. However, autoencoders exhibit some limitations. First, it is very difficult to interpret the underlying mathematics behind them. Second, they can easily overfit. You can refer to [71] for a detailed mathematical description of autoencoders.

### 3.3.2 Deep belief networks

DBNs are types of UPN initiated in 2006 [75]. They contain multiple hidden units in the hidden layers that help to learn higher level representations from input variables. The DBNs comprise restricted Boltzmann machines (RBMs) that deal with higher level features in an unsupervised manner and a feed-forward network that updates the parameters of the model [69]. Nevertheless, units between the hidden layers in DBNs are bidirectional, which makes DBNs differ from other feed-forward neural networks. Additionally, there are no connections between units within the same layers. This difference is accomplished by the RBM layers in the pre-trained phase. The main advantage of DBNs is that they can exploit a large amount of unlabeled data as they work on the principle of layer-wise pre-training. However, the primary cons of DBNs lie in the fact that they are difficult to train. Recently, Refs. [11, 76] have tried to apply DBNs in sentiment analysis. For a detailed mathematical description of DBN, we recommend [71].

### 3.3.3 Generative adversarial networks

The GAN is a kind of UPN invented by Ian Goodfellow at Google brain in 2014 [77]. GAN has discriminative ( $D$ ) and generative ( $G$ ) models that are trained in an unsupervised and competitive fashion. The generative model learns the data distribution of target data while the discriminative model evaluates if examples are from the training set or from the generative model. We illustrate the GAN framework with  $G$  and  $D$  during the training process in Figure 2.

In the original formulation of the GAN problem [77], it is perceived as a game of forging money. In the game,  $G$  plays the role of forging bills while  $D$ , the expert, tries to recognize those forged bills. Therefore, as  $D$  discovers the fake bills,  $G$  improves its skills of producing fake bills accordingly. Following the definition in [78], the GAN game is mathematically defined as follows: let  $x$ ,  $z$  and  $R$  be the characteristic of real bills, fake bills and the metric that measures the real bill, respectively. Therefore, the task of  $D$  is to reduce the quantity of making fake bills  $R(G(z))$  as close as possible to  $R(x)$ , the standard of real bills. Meanwhile,  $G$  also keeps trying to find a way to increase  $R(G(z))$ . Roughly speaking, both models keep depending on each other but do not control each other's parameters. Thus, the game ends when both quantities are in a Nash equilibrium, i.e., when  $D$  has arrived at the optimum point.

In brief, the  $D$  and  $G$  play the minimax game, also called sum-zero-game, which is described in the following [77]:

$$\min_G \max_D R(D, G) = \mathbb{E}_x \sim p_{\text{data}}[\log D(x)] + \mathbb{E}_z \sim p_z[\log(1 - D(G(z)))]. \quad (1)$$

Therefore, as described above,  $D$  wants to maximize the quantity  $R(D, G)$  by making both  $\log D(x) = 1$  and  $D(G(z)) = 0$ . Also,  $G$  wants to minimize the quantity  $R(D, G)$  by making  $D(G(z)) = 1$ .

In sentiment analysis, GAN models are at the inception stage since only a few relevant contributions are found. Li et al. [79] proposed a GAN model for category sentence generation. The generator of the proposed model is a long short-term memory (LSTM) implemented as a reinforcement learning agent while CNN is implemented as a discriminator. Similarly, Vlachostergiou et al. [80] designed a GAN model that is implemented using a denoising autoencoder in order to generate useful representations for different NLP tasks including sentiment analysis.

However, GANs exhibit the limitations that are mainly related to their training process. Li et al. [81] highlighted that the training dynamics of GANs often present two challenges: mode collapse and vanishing gradients. To solve the issues, Li et al. [81] proposed a model that guarantees the GANs to converge. Furthermore, the training of GANs requires finding the Nash equilibrium, which is considered to be less efficient than optimizing the actual objective function [78]. Therefore, this problem is still an open issue for the GANs research community. Moreover, the application of GANs in the NLP domain is still hindered by convergence issues and the problem of dealing with discrete data [82].

### 3.4 Convolutional neural networks

CNNs are types of feed-forward neural networks introduced in 1989 [83]. The application of CNN in the artificial neural network domain was inspired by the process of the animal visual cortex. Each individual neuron of the visual cortex covers a small receptive field, and then receptive fields overlap to visualize the entire object. Thus, receptive fields are considered as filters in CNN. In principle, CNN has three types of layers, namely, the input layer, the feature extraction layers and the classification layer [69]. The input layer takes the raw inputs and produces the embeddings. Next, the feature extraction layers, which include convolution and pooling layers, learn the relevant features. The convolution layer applies filters known as feature detectors to learn the features and produce the feature map. The pooling layer, also known as the dimensional reduction method, is used to extract relevant features, leaving those unnecessary ones. Finally, the features produced by the feature layers are passed to the classification layer, which is made of a fully connected network with a classifier.

The application of CNN in NLP was initiated by Collobert et al. [58]. Since then, CNNs have attained inspiring results in NLP [84, 85], especially in sentiment analysis where the first kickstart was led by [6, 86]. Therefore, the success of CNNs is attributed to the following advantages. First, they have few parameters so that it takes CNNs less time to train compared to other full connected neural networks with the same number of hidden layers [71]. Second, in sentiment analysis, they are good at learning local contextual features using filters. However, in sentiment analysis, CNNs are limited in modeling long-term dependencies because they need to be very deep [87, 88], which leads them to be computationally expensive. Furthermore, their performance is conditioned on a large number of training samples,



which are not always available. Thus, RNNs were introduced as efficient methods that help to model long-term dependencies in sequential input.

### 3.5 Recurrent neural networks

RNNs are a member of feed-forward neural networks that follow the principle proposed by Elman in 1990 [89] to process sequential information. However, RNNs slightly differ from other feed-forward neural networks by directed cycles, which play an important role in the propagation of activation function to the incoming input sequence. Therefore, RNNs are said to have a memory as they can easily remember the state of previous computations. The RNNs consider the output to depend on the previous computation. Thus, RNNs present the following key benefits that have stimulated their suitability for modeling sequential input. Primarily, RNNs have shown the capability of modeling sequences input vectors of any length and long-range dependencies [90]. Secondly, RNNs easily take into account contextual information at each time step while processing sequences data. Although traditional RNNs were proved to get good results by modeling long-range dependencies, they are limited to a certain extent of range between dependencies. Their limit is linked to the gradient vector that increases or decreases proportionally with long-range dependencies. Thus, the problem caused by this increase and decrease is referred to as exploding gradient and vanishing gradient, respectively [91]. Remedy to the problem is the invention of LSTM.

**LSTM.** In 1997, Hochreiter et al. [92] proposed the LSTM to face the issue of vanishing gradient by extending the classic RNN with a gating mechanism. They introduced the forget gate that allows the memory cell to keep information for a long time or throw previous computation results. However, LSTM is accused of having a complex structure. Thus, gated recurrent unit (GRU) was designed as a simplified variant of LSTM.

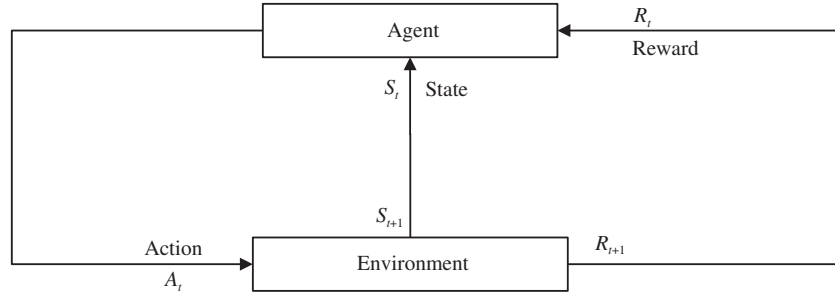
**GRU.** GRU is another variant of RNN, which is similar to LSTM [93]. GRU has two layers unlike LSTM, which has three layers. The first gate named reset gate governs the combination of new input and previous computations. The second gate dubbed update gate determines what information to keep from the previous computations. GRU is considered as a simplified LSTM model and more efficient in terms of computational power compared to LSTM and vanilla RNN.

Modeling the long-range dependencies is not the last step to improve RNNs capability. A recent trend is to allow a model to pick contextual information at every time step. Gratefully, attention mechanism [94] helps to realize it. In 2015, inspired by the human vision capability of focusing on a region of interest, researchers in NLP introduced the attention mechanism. This technique helps to prioritize relevant parts of the given input sequence based on its weighted representation. Moreover, since 2016, the recent surge in sentiment analysis is to apply the cognitive based attention [95,96], which simulates the human reading capability of gaze-fixations on the essential part of the input to be retained. This mentioned form of attention is based on the computation of reading time for every word of the input using eye-tracking movement.

RNNs based models have been remarkably applied in sentiment analysis. For example, Ref. [97] proposed an LSTM, Ref. [98] designed a GRU, Ref. [99] suggested bidirectional LSTM (Bi-LSTM), and Ref. [100] explored attention based bidirectional (Bi-GRU). Although significant progress has been made in improving LSTM and GRU capability, they still have some pitfalls. Initially, it is hard to train them because the memory adds several weights to each node of sequence input during training. Fortunately, GPU has alleviated the problem of training complex models. Furthermore, they are not able to deal with sequential input of arbitrary shapes like a tree. Thus, recursive neural networks allow RNNs to deal with input of arbitrary shape.

### 3.6 Recursive neural networks

In 1996, Goller et al. [101] introduced RvNNs, which is viewed as a generalized version of RNN. RvNNs allow the neural networks to deal with structured inputs of any shape like trees and graphs. In contrast to the original RNN that shares the weights across the whole input sequence, RvNNs allow the weights



**Figure 3** Agent-environment interaction in reinforcement learning framework.

to be distributed across each node of the structured representation of the input. In sentiment analysis, the commonly used variants of RvNN are recursive neural tensor network (RNTN) and semi-supervised recursive autoencoder (RAE).

**RAE** is a variant of RvNN designed in 2011 [102], which applies the same autoencoder to each node of the binary tree. The core purpose of RAE is to model the semantic representation of a sentence by taking into consideration the word order. Thus, while training RAE, the main goal is to minimize the sum of the reconstruction loss at each node of the tree.

**RNTN** was introduced in 2013 [103]. Originally, RNTN was designed for sentiment analysis. Its main intention is to capture the sentiment of a sentence with any length by not only relying on its components but also exploring the order in which words are syntactically grouped. Thus, RNTN carries out this representation based on a tensor-based composition function, which is applied to all nodes of the tree. However, RNTN presents two distinctive features compared to RAE. Firstly, the tree structure is fixed. Secondly, the reconstruction loss in RNTN is ignored. Consequently, the last difference is counted as the main advantage of RNTN over RAE, as the reconstruction loss increases the computational overhead at each node.

RvNNs have shown impressive results in sentiment analysis due to their capability of representing the sequence input in the form of the tree and the ability to represent the context in which a word appears. Furthermore, their ability to learn semantic and syntactic information from the inputs leads them to succeed in sentiment analysis [104, 105].

### 3.7 Deep reinforcement learning

The DRL is an emerging area in the field of DL where the first kick-starting was in 2013 by the DeepMind Technologies group [106]. Generally, DRL integrates the advantage of DL's great insight and that of reinforcement learning (RL)'s decision making and then gains the output control directly from crude input by an end-to-end learning process [107]. Their integration resulted in a substantial breakthrough in a lot of tasks that require the great insight of high-dimension crude inputs and policy control [108]. However, the DRL framework is different from other DL methods for the fact that it does not require supervision. Instead, it only emphasizes the interaction of an agent with its environment and then the agent directly gets a reward signal. Therefore, this is considered as the primary advantage it offers. Consequently, it can be used to solve the problems in which obtaining the target labels is difficult. You can refer to [109], for details of reinforcement learning from its inception.

DRL frameworks are based on Markov decision process (MDP) that helps to train an intelligent agent, which interacts with the environment at each time. The goal of this interaction is to maximize the long-term reward as shown in Figure 3. At the time step  $t$ , the agent receives a high dimension observation from the environment and uses DL to get a specific feature representation of  $S_t \in S$ . Afterwards, the intelligent agent judges adaptive expectations of various actions and maps the current state  $S_t \in S$  to the action  $A_t \in A$  guided by the policy  $\pi(A_t|S_t)$ . The policy is considered as a lookup table of the received states and the actions to be taken for those states. Later on, as a consequence of its actions, the agent receives a new reward  $R_{t+1}$  and achieves a new observation from the environment. DRL methods are systematically

categorized into two main categories: value-based and policy-based DRL methods. The core idea of the value-based DRL method is to approximate the value function by using deep neural networks, which help to get good performance in large-scale discrete action space. The commonly used algorithm in this category is deep Q-network (DQN) [110]. While the policy-based DRL algorithm emphasizes to optimize parametrized policies by targeting the long-term reward. For policy-based DRL, sentiment analysis has witnessed the use of the REINFORCE algorithm [111] for training models with complex structure.

Sentiment analysis heavily relies on the learned representations produced by word embedding methods. However, these models lack a task-specific structure. Therefore, by considering the obtained word and sentence representations as a sequential decision problem, DRLs have shown the promising capability of automatically obtaining an optimized structure, which is useful in NLP, especially in sentiment analysis [112, 113]. Zhang et al. [112] considered the current action of structure discovery to have an effect on the subsequent decisions, which can be considered in the policy gradient method. Like any other classification task, in this setting, the reward is the probability of predicting the correct label to the input sentence based on the obtained structured representation. Training different models with such algorithm have shown improvement in results compared to the supervised methods. Chen et al. [114] enjoyed the beauty of REINFORCE by training a gated multi-modal embedding and LSTM with attention (GME-LSTM(A)) proposed for multi-modal sentiment analysis. In GME-LSTM(A) framework, the controller is implemented as an agent that interacts with the environment by receiving the weights and inputs. Afterwards, it decides to reject or accept the inputs based on the mean absolute error (MAE) received as a reward.

However, despite the advantages offered by reinforcement learning, it presents some disadvantages. First, it is very hard to design a model in which the reward function will be trained [115]. Second, RL models suffer from data inefficiency, which is the high rate of agent-environment interactions [116].

### 3.8 Hybrid deep neural networks

Individual DL models have been extensively used and proved to produce impressive results in NLP [117], especially for sentiment analysis. Therefore, different researchers have tried to combine these approaches to improve the performance of their models by getting the benefits offered by each type. For example, CNNs are well-recognized models in extracting local features. On the other hand, RNNs are well-known to deal with long-range dependencies. Hence, the appealing idea is to integrate them so that the model can extract both types of features. Recently, researchers [12, 118, 119] have proposed various hybrid models to accomplish various tasks in sentiment analysis.

## 4 Real world datasets for sentiment analysis

Getting real-world datasets in academic research is always of great value. Therefore, before we dive into DL approaches application and their performance on different tasks of sentiment analysis, it is worthwhile introducing the most popular real-world datasets that are publicly available and currently used in sentiment analysis literature. The details of each dataset are presented in Table 1. We bring to the readers' attention that user choice means that the splits for the dataset are not specified. Therefore, the user can randomly split as he/she wants.

- IMDB large movie review<sup>5)</sup> [120]. IMDB is a dataset that contains movie reviews and is used in binary sentiment classification.

- IMDB2<sup>6)</sup> [121]. The IMDB2 is a dataset commonly used in document level sentiment analysis. This dataset presents the advantage of providing opinion holder information. Note: the original name of this dataset is IMDB, so to differentiate it from the first IMDB Large Movie review we named it IMDB2.

- Stanford sentiment treebank (SSTb)<sup>7)</sup> [103]. The SSTb dataset consists of movie reviews collected

5) <http://ai.stanford.edu/~amaas/data/sentiment/>.

6) <https://github.com/nihalb/JMARS/tree/master/data>.

7) <https://nlp.stanford.edu/sentiment/>.

**Table 1** Details of the datasets commonly used in sentiment analysis

Dataset	#Samples	#Train	#Dev	#Test	#Classes	Sentiment analysis task	Language
IMDB	50000	25000	–	25000	2	Sentence sentiment classification	English
IMDB2	348415	User choice	User choice	User choice	10	Document sentiment classification	English
SST-5	11855	8544	2210	1101	5	Sentence sentiment classification	English
SST-2	9613	6920	872	1821	2	Sentence sentiment classification	English
Amazon	[10K–143M]	User choice	User choice	User choice	2, 5	Sentence and ABSA	English
SemEval2014-D1	2931	2292	–	639	3	ABSA and ATE	English
SemEval2014-D2	4712	3591	–	1121	3	ABSA and ATE	English
SemEval2017	61873	43601	5988	12284	3	Sentence sentiment classification	English
STS	1.049M	1.048M	–	498	2	Sentence sentiment classification	English
STS-Gold	2034	User choice	User choice	User choice	2	Sentence sentiment classification	English
STS-Gold	58	User choice	User choice	User choice	2	Entity sentiment classification	English
Yelp	Above 1.2M	User choice	User choice	User choice	2, 5	Sentence, document and ABSA	English
HR	24348	User choice	User choice	User choice	2	Sentence sentiment classification	Chinese
MR	10662	8655	961	1046	2	Sentence sentiment classification	English
Sanders	5513	User choice	User choice	User choice	4	Sentence sentiment classification	English
Deutsche Bahn	21824	User choice	User choice	User choice	3	Multi-lingual sentiment	Deutsch
ASTD	10006	User choice	User choice	User choice	5	Multi-lingual sentiment	Arabic
YouTube	47 videos	User choice	–	User choice	3	Multimodal sentiment	English
CMU-MOSI	93 videos/ 2199 utterances	52 videos/ 1284 utterances	10 videos/ 229 utterances	31 videos/ 686 utterances	7	Multimodal sentiment classification	English
CMU-MOSEI	3229 videos/ 22676 utterances	2250 videos/ 16216 utterances	300 videos/ 1835 utterances	679 videos/ 4625 utterances	7	Multimodal sentiment and emotion classification	English

from the Rotten Tomatoes review site. The SSTb dataset is used for binary classification (SST-2) and for fine-grained classification (SST-5) where the reviews are classified into five categories.

- Amazon product review datasets<sup>8)</sup> [122]. The review sentences in these datasets are annotated according to their star number from one to five. Amazon product review datasets are categorized according to the product type with reviews varying from 10261 to 8898041. Moreover, one may choose to use the whole dataset with all product reviews, which includes 142.8 million reviews.
- SemEval-[year]. These datasets are generated annually by the International Workshop on Semantic Evaluation. For example, SemEval2014 task 4<sup>9)</sup> contains user reviews in laptop domain (SemEval2014-D1) and restaurant domain (SemEval2014-D2). The work [123] used 2931 samples for D1 and 4712 samples for D2.
- Stanford Twitter sentiment corpus (STS)<sup>10)</sup> [124]. The STS dataset contains English tweets collected between April and June in 2009. The tweets in this dataset were grouped in positive and negative classes using emoticons as noisy labels.
- STS-Gold dataset for Twitter<sup>11)</sup> [125]. This dataset has been collected as a complement to the STS corpus in which the tweets and entities are labeled individually.
- Yelp challenge datasets<sup>12)</sup>. Yelp datasets contain review texts given by users for products and services available at the Yelp review site. Review texts in these datasets contain the sentiment intensity ranging from one to five stars. Yelp datasets are generated in different rounds, for example, Yelp 2014, Yelp 2015, Yelp 2017 contain 1.2, 1.6, 4.1 million of review sentences, respectively.
- Multimodal corpus of sentiment intensity and subjectivity analysis (CMU-MOSI)<sup>13)</sup> [126]. CMU-MOSI is one of the datasets that are used in multimodal sentiment analysis. It consists of 93 annotated videos collected from YouTube that are divided into 2199 labeled clips/utterances. Each video consists of one speaker who comments on the movie.
- CMU multimodal opinion sentiment and emotion intensity (CMU-MOSEI)<sup>14)</sup> [127]. CMU-MOSEI is the recently largest dataset in multimodal sentiment analysis and emotion recognition. CMU-MOSEI consists of 3229 annotated video clips that are divided into 22676 labeled clips/utterances. The videos are collected from YouTube for 1000 different speakers and cover 250 topics.
- Movie-review (MR) dataset<sup>15)</sup> [128]. MR dataset consists of movie reviews collected from the IMDB reviews website and the sentences are labeled as positive and negative.
- Sanders Twitter sentiment dataset<sup>16)</sup> [129]. This dataset comprises tweets about Google, Twitter, Apple and Microsoft products. The tweets in the dataset are classified into positive, negative, neutral, and irrelevant class.
- German-Deutsche Bahn<sup>17)</sup> [130]. German-Deutsche Bahn dataset consists of 21824 reviews about Deutsche Bahn German public train service.
- Arabic sentiment tweets dataset (ASTD)<sup>18)</sup> [131] consists of 10006 tweets collected from EgyptTrends about political issues. The tweets in ASTD are categorized into positive, negative, mixed, and objective class.
- YouTube<sup>19)</sup> [26]. YouTube is a multimodal sentiment dataset collected on YouTube without considering any specific topic.

---

8) <http://jmcauley.ucsd.edu/data/amazon/>.

9) <http://alt.qcri.org/semeval2014/task4/index.php?id=data-and-tools>.

10) <http://cs.stanford.edu/people/alecmgo/trainingandtestdata.zip>.

11) [https://github.com/pollockj/world\\_mood/tree/master/sts\\_gold\\_v03](https://github.com/pollockj/world_mood/tree/master/sts_gold_v03).

12) <https://www.yelp.com/dataset/challenge>.

13) <https://www.amir-zadeh.com/datasets>.

14) <https://github.com/A2Zadeh/CMU-MultimodalSDK>.

15) <https://www.cs.cornell.edu/people/pabo/movie-review-data/>.

16) [https://github.com/zfz/twitter\\_corpus](https://github.com/zfz/twitter_corpus).

17) <https://sites.google.com/view/germeval2017-absa/>.

18) <http://www.mohamedaly.info/datasets/astd>.

19) <http://projects.ict.usc.edu/youtube/>.

## 5 Deep learning applications on different sentiment analysis tasks

In this section, we summarize the application of DL models on various tasks of sentiment analysis. Our main objective is to depict the current progress of DL application on various sentiment analysis tasks and the results achieved on different real-world datasets.

### 5.1 Document level sentiment analysis

A plethora of research has been conducted to perform sentiment analysis on the document level, which involves finding the polarity of the whole opinionated text. So, in this subsection, we present some representative studies in the domain.

#### 5.1.1 Application

We sort out DL models designed for document level sentiment analysis into five parts: CNN, RNN with attention, RNN with user and product attention, adversarial network and hybrid models.

**CNN based models.** CNN based models have been applied to realize document level sentiment analysis by using different filters to learn local features from the input sentence. Johnson et al. [88] introduced a deep pyramid CNN (DPCNN) model that addresses the issue of long-range dependencies in text documents without increasing the computational cost. The DPCNN model takes input from embeddings, and afterwards convolutional and max-pooling down sampling layers alternate in the form of a pyramid shape. In this way, the max-pooling layers allow a better representation of the long-range dependencies. Later on, Conneau et al. [87] explored the usage of a very deep CNN (VDCNN) that learns the hierarchical representations of the document and learning long-range dependencies. To achieve its objective, VDCNN uses 29 convolutional layers with small convolutions to deal with character level information. The feature maps generated by the convolutional layers are passed to the k-max pooling, which produces fixed dimension features that are fed to the full connected classifier.

**RNN with attention-based models.** Attention mechanism has shown the capability to help RNN to deal with good representations of a document and to capture long-range dependencies at a low computational cost. In the same vein, Yang et al. [132] designed a hierarchical attention network (HAN) model that uses two levels of Bi-GRU with attention to select relevant words and sentences in the construction of the document representation. The HAN model hierarchically constructs the representation of the sentence, and afterwards aggregates the sentence representations to obtain a full document representation sent to the classifier. Similarly, inspired by human reading capability, Long et al. [96] proposed an LSTM model augmented with a cognition-based attention model (CBA), which is trained by cognition grounded eye-tracking data. Initially, the LSTM+CBA model predicts the total reading time of the sentence. Afterwards, to capture relevant words and sentences in the construction of document representation, the attention model constructed from the predicted reading time of the sentence, context, and syntactic features is applied.

**RNN with the user and product attention based models.** Researchers have demonstrated that it is useful to consider opinion holder preference and the product information in sentiment analysis. Chen et al. [97] pioneered the trend by designing a model named neural sentiment and user product attention (NSC+UPA). The NSC+UPA jointly learns sentence and document representations using LSTM. Then, it applies the attention mechanism to the produced representations for prioritizing the most contributing user and product information. Similarly, Dou [13] suggested a model dubbed user and product with deep memory network (UPDMN), which combines LSTM and deep memory model. The LSTM helps the model in the representation of the input document. The model uses the LSTM to construct the document representation. Subsequently, to compute the ratings of each document, the model applies deep memory layers that consist of the attention-based models. In the same manner, Wu et al. [14] proposed an attention LSTM based model named hierarchical user attention and product attention (HUAPA). In contrast to NSC+UPA and UPDMN models, HUAPA separately learns the user and the product representations. And then, it aggregates them to make the final representations. Recently, Amplayo et al. [12] designed



**Table 2** Performance of DL models on document level sentiment classification task

(a) IMDB2 dataset				(b) Yelp_full_2013 dataset			
Ref.	Year	Model	Accuracy (%)	Ref.	Year	Model	Accuracy (%)
[14]	2018	HUAPA	55.0	[14]	2018	HUAPA	68.3
[12]	2018	HCSC	54.2	[132]	2016	HAN	68.2
[97]	2016	NSC+UPA	53.3	[12]	2018	HCSC	65.7
[96]	2017	LSTM+CBA+UPA	52.1	[96]	2017	LSTM+CBA+UPA	65.5
[132]	2016	HAN	49.4	[119]	2015	LSTM-GRNN	65.1
[13]	2017	UPDMN	46.5	[97]	2017	NSC-UPA	65.0
[119]	2015	LSTM-GRNN	45.3	[13]	2017	UPDMN	63.9

(c) Yelp_full_2014 dataset				(d) Yelp_full_2015 dataset			
Ref.	Year	Model	Accuracy (%)	Ref.	Year	Model	Accuracy (%)
[132]	2016	HAN	70.5	[132]	2016	HAN	71.00
[14]	2018	HUAPA	68.6	[88]	2017	DPCNN	69.42
[119]	2015	LSTM-GRNN	67.1	[119]	2015	LSTM-GRNN	67.60
[96]	2017	LSTM+CBA+UPA	66.8	[87]	2017	VDCNN	64.72
[97]	2017	NSC-UPA	66.7				

a hybrid contextualized sentiment classifier (HCSC) that combines a CNN model used to learn short sentences and a Bi-LSTM that deals with long sentences. Certainly, HCSC is in accordance with the previous studies in some way but the main difference lies in a cold start attention model, which helps to tackle the cold start problem caused by user and product information in different reviewed texts.

**Adversarial network models.** Motivated by the theory of domain adaptation where the prediction is accomplished using common features between sources and target domains, various researchers have performed document level sentiment analysis. Ganin et al. [133] designed a model named domain-adversarial neural network (DANN), which is trained with gradient-based optimization directly on the input to a classification network. DANN model trains the classifier using both labeled and unlabeled data from the source domain and target, respectively. Afterwards, as the training progresses, the discriminative features from the source domain and indiscriminative features with regard to the change between domains are promoted. Similarly, Li et al. [134] constructed an adversarial memory network (AMN) model that contains sentiment and domain classifier modules. Both modules are trained together to reduce the sentiment classification error and allow the domain classifier not to separate both domain samples. The attention mechanism is incorporated into the model to help in the selection of the pivots words, which are words useful for sentiment classification and shared between the source and target domains.

**Hybrid model.** To exploit the benefits offered by RNN and CNN in document level sentiment analysis, Tang et al. [119] used a hybrid model that combines gated recurrent neural network (GRNN), CNN and LSTM. This model learns the representation of sentences using CNN or LSTM. At the final step, GRNN learns the semantic and the relationship between sentences to find the overall document sentiment.

### 5.1.2 Performance of DL models on document level sentiment classification

DL approaches have achieved overwhelming results on document level sentiment analysis as shown by Table 2. Based on the results, we observe that the majority and the top performing models on the four datasets (a) IMDB2, (b) Yelp\_full\_2013, (c) Yelp\_full\_2014, and (d) Yelp\_full\_2015 are attention recurrent based models. Thus, this demonstrates the capability of recurrent based models to deal with sequential data like text [90]. In addition, the top performing models are attention based, and this implies the advantages offered by the attention mechanism to prioritize important information and help RNN based models to address long-term dependencies. Furthermore, the benefit of considering user and product information in sentiment analysis is witnessed by the good performance attained by all models that take it into account [12, 96].

## 5.2 Sentence level sentiment classification

In this subsection, we present different DL models that have been suggested on sentence level sentiment classification, which aims to find the class label of an opinionated sentence.

### 5.2.1 Application

DL models proposed to realize sentence level sentiment classification are classified into six parts: UPN, CNN, RvNN, DRL, RNN, and RNN with cognition attention based models.

**UPN based models.** Initially, inspired by the dimensionality reduction capability of autoencoders, Rong et al. [73] introduced an autoencoder based bagging prediction (AEBPA) model, which uses stacked autoencoder with multiple layers as feature learning parts that help get the high-level representation of data. Each stacked auto-encoder pre-trains the network layers with unlabeled data, and finally the network is fine-tuned with the supervised training mechanism. Alike, Zhai et al. [135] designed a model based on a semi-supervised autoencoder by which its loss function is relaxed to the Bregman divergence (SBDAE). The SBDAE solved the problem of task-irrelevant and high dimension. In addition, SBDAE takes into account the label information to guide the feature learning process by allowing the learned representation to be directly associated with the task of the interest. Similarly, Jin et al. [76] suggested a DBN which is extended by the RBM unsupervised training phase. The RBM generates a hidden layer that serves as the input of the single layer feed-forward network, which is trained using the delta rule. Likewise, Jin et al. [11] designed a model that incorporates the word positional and order information to DBN. The constructed DBN comprises the RBM used to train the initial weights in an unsupervised fashion which are transferred to the feed-forward network for backpropagation. Li et al. [79] initiated the use of GANs in sentiment analysis by combining CNN, LSTM, and RL. In the proposed model, the combination of GANs with RL helps to increase the size of the dataset by generating enough sentences for classification. The LSTM implemented as an agent of RL is used as a generator to deal with the sentence structure at each time step whereas the CNN implemented as discriminator distinguishes the category of the sentence generated and provides the reward signal to LSTM. Similarly, Vlachostergiou et al. [80] implemented GANs to learn useful representations in NLP, especially in sentiment analysis. Different from other GANs models, the discriminator of the proposed GANs, which is based on denoising autoencoder, does not perform binary classification of real or non-real data. Rather, it assigns the energy score to each of data received from the generator.

**CNN based models.** CNN based models have been extensively applied to sentence level sentiment analysis pioneered by [6, 86]. Kim [86] suggested a simple, yet efficient CNN model with two channels where each channel consists of a single convolution layer followed by max-pooling over time for sentence level sentiment analysis. In particular, this CNN applies multiple filters that contain kernels of different sizes in order to extract a large number of local features in the sentence. This shallow model gained success, and hence is used by other researchers as the baseline model. Similarly, to deal with short and long sentences using CNN, Kalchbrenner et al. [6] proposed a dynamic CNN (DCNN) model that alternates a wide convolution and dynamic k-max pooling to learn the sentence's semantic structure. Compared to traditional max-pooling, the k-max pooling helps DCNN to deal with the sentence of any length. Following the popularity of CNN, Yin et al. [136] designed a model named semantic lexical-augmented CNN (SCNN) that extends CNN with lexical information of word. The SCNN model learns the sentiment information of the words using the SentiWordNet lexicon and the semantic features of the words using word2vec embedding model.

**RvNN based models.** To deal with input of arbitrary structure, RvNN models have been explored in sentiment analysis. Socher et al. [103] initiated their use by constructing a recursive neural tensor network (RNTN) model that helps to deal with the compositionality, which is a challenging issue in sentiment analysis. The RNTN model learns the compositional vector representation for various phrases and requires the construction of a parse tree for the input sentence, which is the only preprocessing step. Later on, Tai et al. [137] designed a recursive tree-LSTM model that exploits the syntactic properties of a text by modeling the state while considering the whole input sequence and different hidden states of

random child units. Likewise, Kokkinos et al. [138] designed a structured Tree-Bi-GRU coupled with the attention model. The incorporated attention model helps to visualize the relevant representation in the construction of the tree.

**DRL based models.** The success of DL models in sentiment analysis relies on the good representation of a sentence. Thus, different deep DRL models have been suggested to address the issue. Yogatama et al. [113] initiated the task by integrating a tree LSTM with a reinforcement learning model. First, the tree LSTM model learns the sequential representation of the sentence. Afterwards, the RL model learns the tree structure of the represented sentence in order to find the appropriate representation. Subsequently, to find the class label for the obtained sentence representation, the RL applies the reward function, which is the likelihood to predict the correct label on the obtained sentence representation. Similarly, Zhang et al. [112] suggested two models that are based on RL to learn the simplest structured representation of the text. The first model named information distilled LSTM (ID-LSTM) learns the sentence representation by filtering only the needed words in a sentence. The second model dubbed hierarchical structure LSTM (HS-LSTM) learns the representation of sentence from word level to phrase level. Finally, the classifier predicts the sentiment label for each word based on the resulting representations from HS-LSTM or ID-LSTM models.

**RNN based models.** Researchers have tried to perform sentiment analysis by not only relying on the use of plain DL approaches but also integrating with some other components. Correspondingly, Zharmagambetov et al. [47] unified a deep RNN with decision tree (DeepRNN+DT) models. The DeepRNN+DT model is built on top of word2vec pre-trained embedding and trained using the traditional RNN. At last, the random forest classifier assigns the sentiment label to each sentence. Similarly, Mousa et al. [139] explored the use of a generative model named contextual Bi-LSTM with a language model (cBi-LSTM LM), which changes the structure of Bi-LSTM to learn the word's contextual information based on its right and left contexts. With the LM model based on the Bi-LSTM, a separate probability distribution is estimated for every sentiment from the training data. The same work constructed a simple discriminative model based on Bi-LSTM that was also applied on the same dataset. To get better results for sentiment classification, both generative and discriminative models were combined to form a single model cBi+LSTM LM+Bi-LSTM (we dub it CBLSTM+LM) classifier.

**RNN with cognition attention-based model.** Performance in sentiment analysis has not yet reached the desired level. Thus, a new signal in sentiment analysis is to apply the cognition attention-based model. Inspired by the human reading capability, Mishra et al. [95] introduced a hierarchical LSTM based model (Sentiment+PoS+Gaze, we dub it Sent+Gaze) trained by cognition grounded eye-tracking data that predicts overall review text's sentiment. First, the proposed model predicts human gaze behavior, parts-of-speech tagging and word's syntactic information at the word level, and afterwards it fuses the predictions of word level to find the representation to be used so as to determine the overall review's sentiment.

### 5.2.2 Performance of DL models on sentence level sentiment classification

Table 3 shows the inspiring results attained by DL methods on sentence level sentiment classification task. Foremost, the first observation is the remarkable performance of BERT introduced in Subsection 3.2. BERT improved the performance accuracy by 5.4% on (b) SST-2. Thus, this shows its power to produce and use contextualized representations. Furthermore, we observed that the top performing models on the four datasets (a) SST-5, (b) SST-2, (c) IMDB, and (d) MR are RNNs based models. Thus, the results confirm the capability of RNN for modeling sequential input. Moreover, the good performance of Tree-GRU(att) model [138] on both SST-2 and SST-5 datasets can be interpreted in two ways. First, it is a justification for the good capability of the recursive model to represent the input sentence in a tree structure. Thus, this representation helps to learn semantic and syntactic information of input. Second, the good performance can be attributed to the attention mechanism that helps to prioritize contextual information in the formation of the tree. Finally, yet importantly, the results reveal that the new signal cognition based attention in Sent+Gaze [95] is a promising step in sentiment analysis as it outperforms

**Table 3** Performance of DL models on sentence level sentiment classification task

(a) SST-5 dataset				(b) SST-2 dataset			
Ref.	Year	Model	Accuracy (%)	Ref.	Year	Model	Accuracy (%)
[138]	2017	Tree-BiGRU(att)	52.4	[63]	2018	BERT_LARGE	94.9
[137]	2015	Tree-LSTM	51.0	[138]	2017	Tree-BiGRU(att)	89.5
[112]	2018	ID-LSTM	50.0	[86]	2014	CNN	88.1
[112]	2018	HS-LSTM	49.8	[137]	2015	Tree-LSTM	88.0
[86]	2014	CNN	47.4	[136]	2017	SCNN	87.9
[103]	2013	RNTN	45.7	[103]	2013	RNTN	87.6

(c) IMDB dataset				(d) MR dataset			
Ref.	Year	Model	Accuracy (%)	Ref.	Year	Model	Accuracy (%)
[139]	2017	cBLSTM+LM	92.83	[95]	2018	Sent+Gaze	89.0
[96]	2017	LSTM+CBA+LA	90.10	[112]	2018	HS-LSTM	82.1
[47]	2015	DeepRNN+DT	89.90	[112]	2018	ID-LSTM	81.6
[135]	2016	SBDAE	89.59	[86]	2014	CNN	81.5
[95]	2018	Sent+Gaze	89.42				

its counterparts on (d) MR dataset with a great improvement in accuracy of +6.9%, and the result it achieved on (c) IMDB dataset can be compared with others. Therefore, this is a traceable good point. We did not include in the table the performance results achieved by GANs models [79, 80] because they used the subset of existing datasets and different evaluation metrics.

### 5.3 Aspect-based sentiment analysis

The ABSA is the most vital task in sentiment analysis as it provides detailed information on what someone likes or dislikes about an entity or aspect. ABSA encompasses two main subtasks: ATSA and ACSA. ATSA aims to find the sentiment polarity concerning the target entities/aspects that appear in the text while ACSA aims to predict the sentiment label of a given aspect with predefined categories.

#### 5.3.1 Application

We classify the DL models in aspect-based sentiment classification into six categories: attention based models with aspect information, attention-based models with aspect context, RNN with attention memory, RNN with common sense knowledge, CNN, and hybrid models.

**Attention based models with aspect information.** Wang et al. [140] designed an attention-based LSTM with aspect embedding (ATAE-LSTM) model for aspect sentiment analysis. The ATAE-LSTM model takes both the aspect embedding and word input vector as input in order to gain the advantages of aspect information. In this regard, LSTM produces the hidden vectors containing the information from the input aspect and the word input vector. Meanwhile, the attention mechanism helps the ATAE-LSTM model to attend an important part of the sentence when it is given input as a certain aspect that is targeted. Likewise, Tay et al. [141] suggested an aspect fusion LSTM (AF-LSTM(CONV)) model for ABSA that incorporates aspect information into the neural network. Initially, AF-LSTM models the relationship between an aspect and its contextual words with a novel circular convolution of vectors. Afterwards, the resulted vector from the aspect fusion is passed to the attention layer, which produces the weighted representation of the sentence that is sent to the classifier.

**Attention based models with aspect context.** Modeling the association between the target aspect and the context where it appears have also been explored in ABSA. Accordingly, Zhang et al. [142] proved this with a gated recurrent neural network (Bi-GRNN) model suggested for target sentiment analysis. The Bi-GRNN model deals with noisy text and learns the semantic and syntactic information from input tweets. In addition, it models the contextual information where a word appears. Similarly, Yang et al. [5] designed an ABi-LSTM model that leverages the Bi-LSTM with an attention mechanism. The ABi-LSTM models the context where each target appears by processing the sentence in forward and backward

directions. Afterwards, it applies the attention mechanism to prioritize the target information. Thus, ABi-LSTM demonstrated the capability to learn the target aspect and its features independent of any location in the sentence. Likewise, Yang et al. [143] introduced a context entity aspect (CEA) model with a slight difference to the above methods of considering for multi-entity in ABSA. The CEA model performs ABSA by combining the entity and aspect information in a sentence. Thus, to achieve its objective, CEA consists of a triad of layers: interaction layer that deals with mixing the entity and its aspects information, a position layer that exploits the positional information of both aspect and entity, and lastly an LSTM with attention layer that converts the input sentence into context memory. At the final stage, the output of the LSTM with attention is passed to the classifier, which predicts the label of (entity-aspect) combination. In the similar context of using aspect contextual information, He et al. [144] designed an attention-based LSTM model named LSTM+SynATT+TarRep (we call it LSTM+Sy+Tar) that grasps the opportunity of using semantic information of the target term by combining the representations from auto-encoder and attention model. The autoencoder is used to model the target and aspect embeddings information. Meanwhile, an attention mechanism implemented using the dependency parser is applied to model the context of the target and a small subset of its closely syntactical words in a sentence.

**RNN with attention memory model.** Chen et al. [145] introduced a recurrent attention memory (RAM) model for target sentiment analysis. The RAM model integrates Bi-LSTM, GRU and attention mechanism. The Bi-LSTM generates the memory from word2vec or GloVe pre-trained embeddings. Afterwards, the multiple attentions model pays more attention to the more important parts of the sentence where the target appears. Lastly, results from multiple attentions are combined with GRU, and then softmax that predicts the polarity label of the target aspect is applied to the results. RAM model is evidenced to be able to detect the long-term sentiment.

**RNN with commonsense knowledge model.** With the advancement of DL, a new signal of integrating commonsense knowledge with DL models is attracting the community. Consequently, Ma et al. [146] proposed a model named sentic LSTM for ABSA that supplements LSTM with conceptual level information of the target during the training phase. In addition, sentic LSTM implements the target attention that encodes target's information as well as sentence attention that computes the representation of the sentence with respect to aspect and target.

**CNN based model.** To explore the power of CNN based models for ABSA, Xue et al. [30] suggested a gated convolutional network with aspect embedding (GCAE) model. GCAE with help of Gated Tanh-ReLU scrutinizes the important sentiment features with respect to a given aspect or entity. To realize ACSA, the proposed model uses two parallel CNNs whose outputs are combined with the gated unit. Similarly, to perform ATSA the model is extended with third CNN that extracts contextual information of aspect terms. Along the same line of using CNN, Huang et al. [147] introduced two novel CNNs named CNN-parameterized filters (CNN-PF) and CNN-parametrized gate (CNN-PG). The CNN-PF uses CNN to extract information from the sentence and uses the aspect specific features as parameters to the CNN. While CNN-PG uses the aspect specific features like a gate that controls how information is passed to the pooling layer.

**Hybrid model.** In order to achieve better performance in ATSA, Li et al. [123] introduced a hybrid model named TNet that combines Bi-LSTM and CNN, with context-preserving transformation (CPT) layer being inserted between them. The Bi-LSTM takes word embeddings as inputs, and then generates contextualized hidden representations. Afterwards, the CPT layer computes the contextual information of the target-specific features with an attention-based-Bi-LSTM and stores them in memory. Finally, CNN is applied to learn sentiment information.

### 5.3.2 Performance of DL models on aspect-based sentiment analysis

Aspect-based sentiment analysis has attracted a large body of researchers. Hence, a plethora of DL models have been proposed and have achieved good results as depicted in Table 4. Like previous analysis in Subsections 5.1.2 and 5.2.2, there are two main observations. First, all best performing models on the three datasets, i.e., SemEval2014-D1, SemEval2014-D2 and SemEval2015-D2 contain RNN components.



**Table 4** Performance of DL models on ABSA task

(a) ATSA: SemEval2014-D1 dataset				(b) ATSA: SemEval2014-D2 dataset			
Ref.	Year	Model	Accuracy (%)	Ref.	Year	Model	Accuracy (%)
[123]	2018	TNet	76.54	[143]	2018	CEA	80.98
[145]	2017	RAM	74.49	[144]	2018	LSTM+Sy+Tar	80.63
[143]	2018	CEA	72.88	[145]	2017	RAM	80.23
[144]	2018	LSTM+Sy+Tar	71.94	[147]	2018	PF-CNN	79.20
[147]	2018	PF-CNN	70.60	[147]	2018	PG-CNN	78.93
[30]	2018	GCAE	69.14	[30]	2018	GCAE	77.28
[147]	2018	PG-CNN	69.12	[140]	2016	ATAE-LSTM	77.20
[141]	2018	AF-LSTM (CONV)	68.81	[141]	2018	AF-LSTM (CONV)	75.44

(c) ACSA: SemEval2014-D2 dataset				(d) ATSA: SemEval2015-D2 dataset			
Ref.	Year	Model	Accuracy (%)	Ref.	Year	Model	Accuracy (%)
[143]	2018	CEA	84.44	[144]	2018	LSTM+Sy+Tar	81.67
[140]	2016	ATAE-LSTM	84.00	[123]	2018	TNet	80.69
[141]	2018	AF-LSTM (CONV)	81.29	[146]	2018	sentic LSTM	76.47
[30]	2018	GCAE	79.35				

Thus, these results still confirm the power of RNN neural network models. Second, the significance of attention mechanism is manifested in all datasets and tasks of ABSA. This is evidenced by the fact that all models on ABSA are attention-based, except GCAE [30] and PG-CNN/PF-CNN [147] models, which do not contain attention mechanism. Furthermore, TNet [123] has proved the power of combining Bi-LSTM that deals with long-term dependencies and CNN that learns local sentiment information.

#### 5.4 Aspect term extraction

Aspect extraction/ATE is task pioneered by Hu et al. [20] in 2004. This task involves the extraction of aspects/targets in opinionated text and it acts as the prerequisite of ABSA and opinion summarization. Consider the following restaurant review text: “An average restaurant with great food, but poor service”. This review sentence consists of two aspect terms, i.e., “food” and “service” of the restaurant, and hence they need to be extracted. Traditionally, this task was formulated as a sequence labeling task where the conditional random fields (CRF) [148] and linguistic patterns (LP) [20] were the methods mainly applied. Thus, both approaches proved to be inefficient because the good performance of CRF relies on the availability of a large number of features while linguistic patterns are manual features. Hence, DL was introduced as an efficient remedy to these problems. Therefore, in this subsection, we explore the application of DL approaches to the aspect term extraction task.

##### 5.4.1 Application

We classify DL models in aspect term extraction into two categories: hybrid models and RNN with attention based models.

**Hybrid models.** The application of DL approaches for aspect term extraction has been initiated by Poria et al. [17], where a CNN model augmented with the LP was introduced. The proposed model considers embeddings and POS tag features, and then a convolution layer extracts contextual features from the sentence. Afterwards, it learns if the sentence contains the aspect or not. Similarly, Wang et al. [149] proposed a recursive neural conditional random field (RNCRF) for aspect term extraction. The RNCRF model integrates a recursive model based on the dependency tree of the sentence and a conditional random field (RF). The recursive model learns high-level features and the RF maps them in labels, which help to capture the context where the aspect appears. Likewise, Guo et al. [150] extended Bi-LTSM with dependency transmission (DT) between recurrent units for aspect extraction. The proposed model takes the combination of POS tags, character and word-level embeddings as input, at the end of which the outputs of the model are sent to the classifier controlled by CRF layer that ensures valid predictions.



Besides, Xue et al. [151] introduced a multi-task neural network (MTNA) model that considers both ATE and aspect category categorization (ACC) as complementary tasks. Therefore, MTNA takes input form pre-trained word embedding and applies a Bi-LSTM followed by CNN. Finally, for classification in ATE and ACC tasks, the outputs of both CNN and Bi-LSTM are combined. Similarly, to solve the problem of labeled datasets for ATE, which are very scarce, Giannakopoulos et al. [99] suggested a Bi-LSTM with CRF classifier that performs supervised and unsupervised ATE. For a supervised ATE task, the proposed model uses a character-level Bi-LSTM followed by a word-level Bi-LSTM for extracting the structural feature of a sentence. Finally, to perform sequence labeling, the CRF layer utilizes the feature vector produced by the last Bi-LSTM. Besides, to accomplish unsupervised ATE, IOB (inside, outside, beginning) format is used for sequence labeling of different aspects that are present in review text. Lastly, a lexicon is used to determine the polarity of each aspect. Consequently, a new dataset is created on which Bi-LSTM & CRF classifier is applied.

**RNN with attention-based models.** Li et al. [152] suggested a memory interaction network (MIN) model, which consists of two jointly LSTM models that are used to deal with aspects and their related opinion terms in review text, respectively. MIN extends LSTMs with memory networks in order to allow the interactions between the two components. Finally, for the prediction, MIN ensures that aspects are from the sentimental sentence by introducing a third LSTM, which is used to discriminate sentimental and non-sentimental sentences. Similarly, Wang et al. [98] proposed a coupled multi-layer attentions (CMLA) model, which does not require any linguistic pattern for preprocessing. CMLA consists of two GRU attention based models that interactively work together for exchanging features. The first attention GRU extracts the opinion terms while the second deals with aspect terms extraction. On the other hand, Li et al. [153] performed ATE by first generating the opinion summary, which is used as the feature to guide the aspect prediction. And then, to accomplish the task, they proposed a truncated history-attention and selective transformation network (THA-STN) with two components where each one is built on top of LSTM that generates initial word representations. The first component named selective transformation network performs opinion summarization under the guidance of current target aspect. Subsequently, the attention mechanism is applied to generate a weighted sum of the whole opinion. Finally, the second component dubbed truncated history attention performs aspect detection task by combining the history representations and the weighted sum of opinion as features.

#### 5.4.2 Performance of DL models on aspect term extraction

DL approaches have shown good capability in aspect term extraction as demonstrated in Table 5. First, the model sentic LSTM [146] improved the F-score by 4.98% on (c) SemEval2015-D2 dataset. Thus, this result reveals that the integration of commonsense knowledge into the attention based model is a good point and can be applied to improve the result of a given model. Second, the role of attention mechanism is observed since the top performing models are attention based: CMLA [98] on (a) SemEval2014-D1, sentic LSTM [146] on (c) SemEval2015-D2, and THA-STH [154] (d) SemEval2016-D2. Last but not least, the power of recursive neural networks is shown on (b) SemEval2014-D2 dataset where the model RNCRF [149] leads since 2016.

### 5.5 Emotion detection

Emotion detection is the task of modeling a person's emotion state that may be anger, disgust, love, joy, optimism, pessimism, confusion, sadness, surprise, and trust. Thus, emotion detection is considered as a supervised multi-label classification problem. It was introduced in 2005 by Alm et al. [23] using supervised machine learning. However, traditional approaches to emotion detection including lexicon based [155] are no longer suitable for today's emotion states expressed in different language styles. Therefore, in this subsection we scrutinize the application of DL approaches as a solution to overcome the limitations of traditional approaches.

**Table 5** Performance of DL models on aspect term extraction task

(a) SemEval2014-D1 dataset				(b) SemEval2014-D2 dataset			
Ref.	Year	Model	F-measure (%)	Ref.	Year	Model	F-measure (%)
[98]	2017	CMLA	85.29	[149]	2016	RNCRF	89.33
[17]	2016	CNN-LP	82.32	[99]	2017	Bi-LSTM-CRF-unsup	88.00
[150]	2018	Bi-LSTM-DT	80.22	[17]	2016	CNN-LP	87.17
[153]	2018	THA-STN	79.52	[150]	2018	Bi-LSTM-DT	85.97
[149]	2016	RNCRF	78.42	[153]	2018	THA-STN	85.61
[99]	2017	Bi-LSTM-CRF-sup	77.96	[99]	2017	Bi-LSTM-CRF-sup	84.12
[152]	2017	MIN	77.58	[151]	2017	MTNA	83.65
[99]	2017	Bi-LSTM-CRF-unsup	73.04	[98]	2017	CMLA	77.80

(c) SemEval2015-D2 dataset				(d) SemEval2016-D2 dataset			
Ref.	Year	Model	F-measure (%)	Ref.	Year	Model	F-measure (%)
[146]	2018	sentic LSTM	76.44	[153]	2018	THA-STN	73.61
[153]	2018	THA-STN	71.46	[151]	2017	MTNA	72.95
[98]	2017	CMLA	70.73				
[151]	2017	MTNA	67.73				

### 5.5.1 Application

We categorize DL models that have been proposed for emotion detection task into three parts: attention, CNN, RNN based models.

**Attention based models.** Felbo et al. [156] suggested an attention Bi-LSTM model named DeepMoji. DeepMoji extends the distance supervision by millions of emojis to classify the emotion, sentiment and detect the sarcasm in the text content. Specifically, emojis are used as noisy labels to pretrain models in absence of labeled datasets. In the same vein, Kim et al. [157] suggested attentive CNN for emotion classification. Initially, to deal with limited tweets data, a preprocessing step is applied to map each emoji in the data to phrase/word that represents it. Afterwards, a self-attention model is used to generate the weighted representation of the sentence by which a CNN is applied to extract various features for emotion classification. Similarly, Yu et al. [158] suggested an attention-based transfer learning model named dual attention transfer network (DATN). DATN uses sentiment classification as a way of improving emotion classification. Specifically, the DATN model implements a shared LSTM and target-specific LSTM. The shared LSTM learns sentiment features shared between sentiment and emotion classification tasks whereas the target-specific LSTM layer deals with specific emotion features that are necessary for the emotion classification task.

**CNN based model.** Analyzing emotion in Chinese microblogs has not been left behind. Thus, Wang et al. [159] performed emotion classification for Chinese microblogs by supplementing the skip-gram language model with a CNN trained to perform multi-label classification of emotions. The proposed model presents the advantage of detecting multiple emotions that may be present in a single sentence.

**RNN based models.** Abdul-Mageed et al. [160] extended the distance supervision with emotion hashtags to generate the dataset labels to be used for emotion classification. Afterwards, they developed a GRU based model to learn the representations that are necessary for the emotion classification task. However, different from above-discussed models that consider single genre emotion classification, a multitasking GRU based model named joint multi-task emotion (JMTE) was proposed by Tafreshi et al. [161]. JMTE model helps to deal with multi-genres emotion classification for the following genres: blog posts, news headlines, movie reviews, and tweets. JMTE contains two branches: the first deals with emotion tweets while the second handles the remaining genres. Thus, the two branches share emotion cues, semantic, and syntactic emotion features among the genres. Finally, each branch performs a specific genre classification.

**Table 6** Performance of DL models on emotion detection task

(a) SemEval-18 dataset				(c) TweetEN dataset			
Ref.	Year	Model	Accuracy (%)	Ref.	Year	Model	Accuracy (%)
[158]	2018	DATN2	59.7	[161]	2018	JMTE	80.0
[158]	2018	DATN1	58.2				
[157]	2018	ACNN	57.4				
(b) Ren-CECps dataset				(d) BLG+HLN dataset			
Ref.	Year	Model	Accuracy (%)	Ref.	Year	Model	Accuracy (%)
[158]	2018	DATN2	45.7	[161]	2018	JMTE	84.0
[158]	2018	DATN1	39.3				
				(e) MOV dataset			
Ref.	Year	Model	Accuracy (%)	Ref.	Year	Model	Accuracy (%)
				[161]	2018	JMTE	92.0

### 5.5.2 Performance of DL models on emotion detection

DL models have achieved incredible results on emotion detection task as shown in Table 6. In view of the results, we observe that the application of DL to this task is on the onset stage. Therefore, we only show the results achieved on datasets under consideration.

## 5.6 Multi-lingual sentiment analysis

With the advent of Web 2.0 tools, users often express their ideas in different languages. However, resource-poor languages have been left behind in benefiting sentiment analysis. Therefore, multi-lingual sentiment analysis debut in 2009 [25] as a task of sentiment analysis, which helps to classify opinionated text in different languages. Traditionally, the performance of the different models on the multi-lingual sentiment analysis task relies on the performance of machine translation, which is proved to be a headache task. Recently, DL approaches have contributed to the boom of the task as they automatically learn features. Therefore, in this subsection, we survey DL approaches that have recently been proposed for this task.

### 5.6.1 Application

We arrange the DL models in multi-lingual sentiment analysis into four categories: UPN, CNN, RNN, and attention based models.

**UPN based models.** Zhou et al. [74] designed a denoising auto-encoder (CLSCDA) model for cross-lingual sentiment analysis. The CLSCDA model explored the use of rich English resources for Chinese sentiment analysis by training English and Chinese classifiers on English labeled reviews and English-to-Chinese labeled reviews, respectively. For testing, the English classifier and Chinese classifier are tested on English-to-Chinese and Chinese reviews, respectively. Finally, to make the prediction, the results from both classifiers are combined. In the same way, Chen et al. [162] proposed an adversarial deep averaging network (ADAN) model to realize cross-lingual sentiment analysis. ADAN comprises sentiment classifier and adversarial language identification scorer components that are concurrently trained. Both parts take input from a shared feature extractor. ADAN is trained with labeled SOURCE text data from English and unlabeled TARGET text data from Arabic and Chinese.

**CNN based models.** Researchers have investigated the use of CNN models to perform multi-lingual sentiment analysis. Consequently, Ruder et al. [163] proposed a CNN model for multi-lingual aspect extraction and aspect sentiment analysis. The model takes embeddings combined with aspect vectors as input. Afterwards, the convolutional layer applies the filters, and then the max-pooling layer is used to learn the desired features for aspect sentiment analysis. Similarly, Becker et al. [164] formulated a CNN that uses the character-based embedding and convolutional layers to learn features from English, Spanish, Portuguese and German. Consequently, the proposed model exhibited the capability of learning important features from all the four languages used during training at one pass, because it does not require the translation process. Besides, Attia et al. [165] designed a language-independent CNN model

**Table 7** Performance of different deep learning approaches at multi-lingual sentiment classification task

Ref.	Year	Model	Dataset	Accuracy (%)
[74]	2014	CLSCDA	NLP-CC	80.02
[162]	2016	ADAN	Arabic	55.33
[162]	2016	ADAN	Chinese	42.49
[165]	2018	CNN	Sanders	79.57
[165]	2018	CNN	Deutsche Bahn	75.45
[165]	2018	CNN	ASTD	65.58
[168]	2018	ACN	Twitter	81.90
[168]	2018	ACN	Weibo	82.43
[167]	2018	BiLSTM	Hindi	72.42

for sentiment multi-lingual analysis. The suggested model presents the advantage of not relying on any language construct features like syntactic features, ontologies, and dictionaries.

**RNN based models.** RNN based models have also been explored for addressing the problem of low resources languages. Can et al. [166] proposed a reusable GRU model for sentiment analysis. The proposed GRU is trained using an English dataset and then the weights are saved. Afterwards, the model is evaluated using the English translated review texts from languages with limited resources. Likewise, Akhtar et al. [167] designed a BiLSM based approach that performs aspect sentiment analysis in resource-poor languages. Initially, they trained a bilingual word embedding and then solves the problem of data sparsity in resource-poor word embedding by supplementing it with word embeddings, which are created from resource-rich languages like English. Eventually, the BiLSM model is applied to learn different features necessary for classification.

**Attention based model.** Wang et al. [168] investigated the use of the attention based CNN (ACNN) model to capture language-specific features. The ACNN is incorporated into the adversarial framework to deal with multi-lingual features. Thus, the model produces good results in bridging English tweets and Chinese Weibo messages.

### 5.6.2 Performance of DL models on multi-lingual sentiment analysis

As demonstrated in Table 7, DL models have produced good results on multi-lingual sentiment analysis. Foremost, we cannot compare the results on this task because there are no different models that have been proposed on a common dataset. A significant observation is that most of the models [74, 162, 168] on this task follow the principle of the unsupervised network for training a multi-lingual classifier. In brief, DL models helped to reduce the burden of translation that was required to perform multi-lingual sentiment analysis. And this needs to be appreciated. However, there are few publicly available datasets for this task. Therefore, we encourage researchers to share the datasets considered in resource-poor languages sentiment analysis.

## 5.7 Multi-modal sentiment analysis

The surge of social media use has brought a high demand for multi-modal sentiment analysis [26] posed in 2011. This task aims at detecting the attitude of the speaker with the help of his verbal and non-verbal behaviors. By virtual of DL, the task becomes easier as different features of modalities are automatically learned with hidden layers of the network. Thus, in this subsection, we present the application of DL to multi-modal sentiment analysis.

### 5.7.1 Application

DL models for multi-modal sentiment analysis are sorted into four categories: CNN, RNN, RNN with attention and DRL based models.

**CNN based model.** Cambria et al. [169] proposed a model that extracts utterances' visual and textual features using CNN and applied openSMILE to learn acoustic features. Afterwards, the three

**Table 8** Performance of DL models on multi-modal sentiment analysis task

(a) CMU-MOSI dataset				(b) CMU-MOSIE dataset			
Ref.	Year	Model	Accuracy (%)	Ref.	Year	Model	Accuracy (%)
[100]	2018	MMMU-BA	82.3	[100]	2018	MMMU-BA	79.80
[170]	2017	bC-LSTM	80.3	[127]	2018	Grah-MFN	76.90
[169]	2017	CNN	78.8				
[171]	2017	TFN	77.1				
[172]	2018	MARN	77.1				
[114]	2017	GME-LSTM(A)	76.5				
(c) IEMOCAP dataset							
Ref.	Year	Model	Accuracy (%)				
[170]	2017	bC-LSTM	76.10				
[169]	2017	CNN	71.59				
(d) YouTube dataset							
Ref.	Year	Model	Accuracy (%)				
[172]	2018	MARN	54.2				
[171]	2017	TFN	47.1				

kinds of feature are fused and then sent to SVM for the final classification of utterance.

**RNN based models.** To exploit contextual features among utterances that make a video, Poria et al. [170] designed a bidirectional contextual LSTM (bC-LSTM) model for multi-modal sentiment classification. bC-LSTM extracts visual features using 3D-CNN. The textual and audio features are extracted using CNN and openSMILE, respectively. Thus, the model considers the order and contextual information between utterance and its surroundings. Likewise, Zahed et al. [171] proposed a tensor fusion network (TFN) model that learns the intra and inter-modalities features in an end-to-end manner. The TFN learns textual, visual and acoustic features using LSTM, FACET tool and COVAREP tool, respectively. Afterwards, instead of directly fusing the features, the model applies a tensor fusion layer that learns the dynamics of each of them explicitly. Finally, each utterance vector with all features is sent to the classifier. Similarly, Zahed et al. [127] proposed a graph memory fusion network (graph-MFN) that learns different features of utterances by applying three parallel LSTMs for visual, audio and acoustic modalities. Subsequently, a dynamic fusion graph (DFG) is applied to learn the importance of each modality and fuses their features for final prediction.

**RNN with attention-based models.** Understanding human communication always requires modeling different modalities and their relationship. To address the issue, Zahed et al. [172] proposed a model dubbed multi-attention recurrent network (MARN), which embodies two components: long-short term hybrid memory (LSTHM) and multi-attention block (MAB). LSTHM learns the view specific within modalities while MAB models the cross-view dynamics between modalities at each time step of recurrent. Similarly, Ghosaly et al. [100] designed an approach named multi-modal multi-utterance-Bi-modal attention (MMMU-BA) that consists of three parallel Bi-GRUs for learning features of video, audio and text, respectively. Afterwards, the attention mechanism is applied to the outputs of each Bi-GRU to prioritize contextual utterances, which correlate to the target utterance.

**DRL based model.** Chen et al. [114] constructed a model titled GME-LSTM(A), which integrates gated multi-modal embedding (GME) and LSTM with temporal attention (LSTM(A)). The GME performs multi-modal fusion by examining the multi-modals at each step and the attention mechanism helps to attend important modalities for the sentiment classification. The GME-LSTM(A) model is trained using reinforcement learning.

### 5.7.2 Performance of DL models on multi-modal sentiment analysis

DL models have shown the incredible capability of dealing with multi-modal sentiment analysis, as depicted in Table 8. First of all, we acknowledge Zahed et al. [127] for creating a new large dataset for multi-modal sentiment analysis. In view of the results, we observe that the top scorer models [100, 172] on (a) CMU-MOSI, (b) CMU-MOSIE, and (d) YouTube datasets are RNN with attention-based models. Thus, the results are in strong agreement with previous studies, which have shown that RNNs are good models for dealing with sequential inputs. Furthermore, the attention mechanism incorporated into the

models helps to prioritize important utterances. In a nutshell, the application of DL on multi-modal sentiment analysis has shown promising results. However, it is still at the baby stage on this task, so more interdisciplinary talents needs to participate in it.

## 5.8 Opinion summarization

Online users generate a huge volume of review texts that are useful to business organization and individual customers. Unfortunately, it is an overwhelming task to find the insight hidden in the reviews by reading through thousands of review texts. To remedy the problem, opinion summarization, also known as aspect-based opinion summarization, was introduced in 2004 [20]. Opinion summarization focuses on generating quantitative concise summaries of a given aspect or entity and related sentiment from a reviewed text. Therefore, in this subsection, we review the application of DL in opinion summarization task.

### 5.8.1 Application

We group DL models applied in opinion summarization task into four categories: CNN, RNN with attention, DRL, and hybrid models.

**CNN based models.** Wu et al. [173] proposed a multitask CNN (MCNN) model for aspect-based opinion summarization. MCNN encompasses multi-channel CNN and single channel CNN. Multi-channel CNN handles the mapping between extracted aspects with respect to their sentiments whereas the single channel CNN learns the necessary features to perform the sentiment classification. Afterwards, the aggregator generates a report containing each aspect and its counts of positive and negative sentiment sentences. Likewise, Li et al. [174] tackle opinion summarization in Chinese microblogs using a CNN and TextRank (TR) + maximal marginal relevance (MMR). CNN automatically learns the representative features from the input text. Afterwards, the outputs of CNN are sent to the TR algorithm that constructs the feature vector graph, and finally, the MMR selects the representative features to be used while constructing the abstract of the whole message.

**RNN with attention-based models.** Wang et al. [175] suggested an attention-based Bi-LSTM model to generate the abstract of review text related to one topic. The proposed model takes multiple input text units and encodes their representations using the attention mechanism, which selects important words in the context of a particular topic. Finally, the importance estimation model produces the outputs importance scores for each text unit, which are used to form the abstract summary. Likewise, Yang et al. [176] designed an approach for opinion summarization that works in the cross-domain setting. Initially, the model classifies the source and target domains by applying an LSTM. Then, a weakly supervised LDA is applied to learn the representations, which are useful for extracting the aspect of a given domain together with its sentiment information. Finally, the attention mechanism is applied to the produced representations. The attention mechanism discovers important aspects together with their corresponding sentiment information that are necessary to the abstractive summary.

**DRL based model.** Yang et al. [177] proposed an end-to-end model trained with the RL algorithm to deal with abstract review summarization. The proposed model consists of encoder and decoder parts. The encoder with multiple attentions learns the representations of context words, sentiment words, and aspects words. Later on, the decoder with attention fusion produces a summary of different aspects and their corresponding sentiment statistics.

**Hybrid model.** Angelidis et al. [178] introduced a neural network model that works in the following manner. At first, with the help of attention encoder, a multi-seed aspect extractor predicts the aspect at the segment level. Afterwards, a CNN is applied to encode the segments' sentiments and produces the outputs, which are used by the attention based GRU in order to produce the final prediction of the document. Finally, the summary is generated by taking the opinions with the highest score for each aspect.



**Table 9** Performance of different deep learning approaches at opinion summarization task

Ref.	Year	Model	Dataset	Accuracy (%)	ROUGE1 (%)	BLEU (%)
[173]	2016	MCNN+w2v	ASM	84.10	–	–
[174]	2016	CNN	COAE2014	86.01	–	–
[177]	2018	MARS	Amazon	–	84.13	–
[176]	2018	CASAS	Amazon-el	–	63.55	–
[178]	2018	MILNET	OPOSUM	–	44.10	–
[175]	2016	AT+Bi-LSTM	RottenTomatoes	–	–	24.88

### 5.8.2 Performance of DL models on opinion summarization task

We illustrate in Table 9 the performance achieved by DL models for opinion summarization. However, we cannot compare the results on this task because the experiments were conducted on different datasets and they used different evaluation metrics. We observed from the results that the application of DL to this task is on the onset stage, and therefore we hope that it will attract more researchers' attention.

## 5.9 Opinion spam detection

Recently, online products and services reviews have become the main source of information, which guides consumers and business organizations in making any decision. However, some kinds of reviews are harmful because they are forged. Therefore, in 2008, researchers led by Jindal et al. [24] have paid attention to such types of reviews known as opinion spam/deceptive opinion spam, which is written to sound authentic, because users cannot easily judge their truthfulness. Therefore, in this subsection, we discuss several applications of DL approaches as a solution for detecting deceptive opinion spam.

### 5.9.1 Application

DL models for opinion spam detection are grouped into CNN based and hybrid models.

**CNN based models.** Zhao et al. [179] suggested a CNN model for deceptive opinion spam detection. The proposed model extended the CNN with word order preserving pooling layer, which allows to keep the order of words in a sentence while analyzing deceptive opinion spam. Likewise, Li et al. [180] analyzed deceptive spam by applying two consecutive CNNs for sentence representation and document representation, respectively. To compute the final representation of the document, each sentence is associated with its importance weight so that only contributing sentences are taken into consideration. In addition, in order to make the prediction, the document representation is supplemented with first-person pronouns and POS tags.

**Hybrid models.** Researchers have tried to combine different models in order to gain the advantages offered by each one. Thus, Ren et al. [181] combined CNN and Bi-GRU with attention mechanism. The model introduced the attention mechanism for detecting and prioritizing important cues of the opinion spam, which help to form the document representation. Likewise, Ren et al. [154] suggested a model that aggregates CNN and Bi-GRU for document level opinion spam detection. The representations of different sentences that make a document are fed to CNN, which produces the sentence features, and then the Bi-GRU learns them to make the final representation to classify the document as spam or truth. Similarly, Zhang et al. [182] designed a recurrent convolution network (RCNN) model that learns the contextual information of each word in review text. Specifically, the model represents each word using its deceptive and truthful contexts, and word embedding. After modeling contextual information of each word, the maxi-pooling is applied to select the maximum features to represent the whole review document.

### 5.9.2 Performance of deep learning models on opinion spam detection task

DL models have performed well in opinion spam detection task as demonstrated in Table 10. Foremost, the observation is that the application of DL to this task is at the infancy stage as we only found five works. Therefore, more importance should be attached to the application of DL models to this

**Table 10** Performance of DL models on opinion spam detection task Hotel-Restaurant-Doctor spam review dataset

Ref.	Year	Model	Accuracy (%)
[182]	2018	DRI-RCNN	86.01
[181]	2016	Bi-GRNN(att)	84.10
[154]	2017	Bi-GRNN	83.60
[180]	2017	SWNN	80.10

task. Furthermore, the results reveal the power of hybrid models, as best performing approaches are all hybrid [154, 181, 182].

## 6 Current problems and future directions

Sentiment analysis helps to extract the hidden insight into the user-generated data. So far, researchers have suggested a large number of deep learning approaches for sentiment analysis that performed well in a large number of tasks. As deep learning research grows, we expect to see this trend to continue with the best models design. However, sentiment analysis has not yet reached the limit. Thus, in the subsequent subsections, we talk about unaddressed tasks, challenges to be addressed and suggestions to improve the performance of existing models.

### 6.1 Unaddressed sentiment analysis tasks

Foremost, Subsection 2.2 highlights various sentiment analysis tasks. However, some of these tasks have not yet been addressed using deep learning approaches. Those tasks include opinion holder extraction and classification as well as time extraction and standardization. Therefore, we hypothesize that an obvious solution is to use different models that have been applied to other domains for NER. You can refer to the recent review of NER using deep learning [183]. Moreover, an important observation is that there is an imbalance of deep learning application between various sentiment analysis tasks. Document level and sentence level sentiment analysis have dominated the rest. Hence, we recommend to keep forward as DL models have proved to gain inspiring results for all tasks including those that are dominated.

### 6.2 Challenges to the future progress

**Dynamic sentiment analysis and tracking.** Currently, most of the deep learning models suggested performing sentiment analysis by assuming that the text exhibits a static nature. However, it is still a challenge to deal with dynamic sentiment analysis and tracking, which involves the dynamic nature of streaming data like in social networks. Let us take an example of a Twitter conversation that involves many users talking about elections. This kind of situation raises the following challenges to sentiment analysis. First, the most challenging issue is the fast-paced change in the dataset at each time step. Second, vocabularies used by users can change over time. Third, the number of users that are involved in the conversation and their preferences can change over time. Therefore, performing sentiment analysis to this sort of situation is still an open research problem. To our knowledge, no deep learning model has been proposed to address this topic. Thus, owing to the capability of deep learning models, especially self-attention mechanism and recurrent neural network of modeling sequential input, we conjecture that modeling social networks with self-attention mechanism or recurrent neural networks by combining the sentiment links and content will help to deal with this situation.

**Sentiment analysis for heterogeneous information.** Dealing with the heterogeneity of information and users in an online social network requires careful analysis and proper methods. Moreover, the interaction between these two components in a social network is still a challenging issue because it requires other additional information like modeling the dependency among users. For instance, modeling and predicting the sentiment information that will follow a president's tweet requires even capturing the influence of the speaker. Therefore, the same as the above issue, a possible solution is to combine

the textual and network information by applying deep learning approaches like a recurrent neural network, convolutional neural network. Recently, attempts have been made in the network embedding field to predict the social influence using a convolution neural network with attention mechanism [184] and sentiment link prediction using auto-encoder [185].

**Language structure.** Thankfully, deep learning models have solved the problem of emotion analysis based on emojis [156, 157]. However, an existing problem is to consider the nature of the language structure, which includes slangs that are used on social media. Let's take an example of the nature of language which is used to write a comment or a message that evolves over time like certain slangs and writing style used on Tweeter, i.e., "b4" means "before" and "kk" means "cool cool". Thus, we speculate that a straightforward solution is to follow the strategies adopted in [156, 157]. Different common slangs can be collected and then used to generate the dataset labels.

### 6.3 Suggestions for application of deep learning to sentiment analysis

Trending deep learning methods such as deep reinforcement learning, generative adversarial networks are at the inception stage in sentiment analysis. Therefore, further research should focus on these methods as they have shown to be promising solutions for complex issues. Precisely, deep reinforcement learning can solve the problem of resource-poor languages because it does not need labeled datasets.

Moreover, GANs can be used to bridge the gap between resource-poor languages and resource-rich languages like English. GANs for this task can be used by considering the resource-rich language data for training the generator and resource-poor languages data for training the discriminator. Furthermore, the new signals, including the tradition attention mechanism [94], self-attention mechanism [64], cognition-based attention models [95, 96], and common sense knowledge [146], have shown the capability of increasing the performance of models, into which these new signals are incorporated. Therefore, we recommend researchers to focus on their use. Beyond using new signals, most of them have proved to increase the performance of the models when they are coupled with RNN based models. Therefore, future research should concentrate on them when they are coupled with other models than RNN based models. For example, Kim et al. [157] took a leading step by combining CNN and self-attention model. Further, the self-attention model [64] has alleviated the issue as it can be used as a standalone model or it can be coupled with other network models. Moreover, we recommend researchers to use sentiment-specific word embedding models [65–67] by requesting the source-code to specific authors we cited. Last but not least, researchers are advised to grasp the advantages offered by BERT [63], which is the recent trending language model that produces contextualized representation learning.

## 7 Conclusion

Sentiment analysis using deep learning approaches has attracted a large number of researchers. Hence, a plethora of deep learning models have been proposed and proved to give good results on various tasks of sentiment analysis. The success of the mentioned approaches is attributed to their capability of automatic feature learning and the success of word embedding models. Therefore, in this study, we first present the background of sentiment analysis, including its applications, levels, and tasks of sentiment analysis. Likewise, we give an introduction of traditional approaches for sentiment analysis and their drawbacks. We mainly discuss deep learning approaches and their applications on different tasks of sentiment analysis. Those tasks involve document level, sentence level, ABSA, multi-lingual, and multi-modal sentiment analysis. Moreover, we address aspect term extraction, opinion summarization, emotion detection, and opinion spam detection tasks in our contribution. Beyond exploring their applications, we provide the performance analysis of the results they achieved on real-world datasets for each task. Finally, we highlight current issues that need to be addressed and provide suggestions for improvement including using new signals, new models like GANs, DRL models, usage of sentiment-specific word embedding models and BERT language model.

**Acknowledgements** This work was supported by National Key Research and Development Program of China (Grant Nos. 2016YFB0800402, 2016QY01W0202), National Natural Science Foundation of China (Grant Nos. U1836204, 61572221, 61433006, U1401258, 61572222, 61502185), Major Projects of the National Social Science Foundation (Grant No. 16ZDA092), and Guangxi High Level 1043 Innovation Team in Higher Education Institutions Innovation Team of ASEAN Digital Cloud Big Data Security and Mining Technology.

## References

- 1 Pozzi F A, Fersini E, Messina E, et al. Challenges of sentiment analysis in social networks: an overview. *Sentiment Anal Soc Netw*, 2017, 1: 1–11
- 2 Liu B. *Sentiment Analysis and Opinion Mining*. Williston: Morgan & Claypool, 2012
- 3 Pang B, Lee L. Opinion mining and sentiment analysis. *FNT Inf Retrieval*, 2008, 2: 1–135
- 4 Liu B. *Sentiment Analysis: Mining Opinions, Sentiments, and Emotions*. Cambridge: Cambridge University Press, 2015
- 5 Yang M, Tu W T, Wang J X, et al. Attention-based LSTM for target-dependent sentiment classification. In: *Proceedings of the 31st AAAI Conference on Artificial Intelligence*, California, 2017. 5013–5014
- 6 Kalchbrenner N, Grefenstette E, Blunsom P. A convolutional neural network for modeling sentences. In: *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics*, Maryland, 2014. 655–665
- 7 Stone P J, Bales R F, Namenwirth J Z, et al. The general inquirer: a computer system for content analysis and retrieval based on the sentence as a unit of information. *Syst Res*, 2007, 7: 484–498
- 8 Pang B, Lee L, Vaithyanathan S. Thumbs up? sentiment classification using machine learning techniques. In: *Proceedings of Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Philadelphia, 2002. 79–86
- 9 Zhang L, Wang S, Liu B. Deep learning for sentiment analysis: a survey. *WIREs Data Min Knowl Discov*, 2018, 8: 1–25
- 10 Rojas-Barahona L M. Deep learning for sentiment analysis. *Lang Linguist Compass*, 2016, 10: 701–719
- 11 Jin Y, Zhang H, Du D L. Incorporating positional information into deep belief networks for sentiment classification. In: *Proceedings of Industrial Conference on Data Mining*, 2017. 1–15
- 12 Amplayo R K, Kim J, Sung S, et al. Cold-start aware user and product attention for sentiment classification. In: *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics*, Melbourne, 2018. 2535–2544
- 13 Dou Z-Y. Capturing user and product information for document level sentiment analysis with deep memory network. In: *Proceedings of Conference on Empirical Methods in Natural Language Processing*, Copenhagen, 2017. 521–526
- 14 Wu Z, Dai X-Y, Yin C, et al. Improving review representations with user attention and product attention for sentiment classification. In: *Proceedings of the 32nd AAAI Conference on Artificial Intelligence*, New Orleans, 2018. 5989–5995
- 15 Salinca A. Convolutional neural networks for sentiment classification on business reviews. In: *Proceedings of IJCAI Workshop on Semantic Machine Learning*, Melbourne, 2017
- 16 Alashri S, Kandala S S, Bajaj V, et al. An analysis of sentiments on facebook during the 2016 U.S. presidential election. In: *Proceedings of IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, San Francisco, 2016. 795–802
- 17 Poria S, Cambria E, Gelbukh A. Aspect extraction for opinion mining with a deep convolutional neural network. *Knowl-Based Syst*, 2016, 108: 42–49
- 18 Wang Z Q, Zhang Y. Opinion recommendation using neural memory model. In: *Proceedings of Conference on Empirical Methods in Natural Language Processing*, Copenhagen, 2017. 1626–1637
- 19 Mooney R J, Bunescu R. Mining knowledge from text using information extraction. *SIGKDD Explor Newsl*, 2005, 7: 3–10
- 20 Hu M Q, Liu B. Mining and summarizing customer reviews. In: *Proceedings of the 10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Seattle, 2004. 168–177
- 21 Turney P D. Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews. In: *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, Philadelphia, 2002. 417–424
- 22 Kim S M, Hovy E. Determining the sentiment of opinions. In: *Proceedings of the 20th International Conference on Computational Linguistics (COLING)*, Geneva, 2004
- 23 Alm C O, Roth D, Sproat R. Emotions from text: machine learning for text-based emotion prediction. In: *Proceedings of Conference on Human Language Technology and Empirical Methods in Natural Language Processing (HLT/EMNLP)*, Vancouver, 2005. 579–586
- 24 Jindal N, Liu B. Opinion spam and analysis. In: *Proceedings of Conference on Web Search and Web Data Mining (WSDM)*, Palo Alto, 2008. 219–230
- 25 Boiy E, Moens M F. A machine learning approach to sentiment analysis in multilingual web texts. *Inf Retrieval*, 2009, 12: 526–558
- 26 Morency L-P, Mihalcea R. Towards multimodal sentiment analysis: harvesting opinions from the web. In: *Proceedings of International Conference on Multimodal Interfaces*, Alicante, 2011. 169–176
- 27 He Y L, Lin C H, Gao W, et al. Tracking sentiment and topic dynamics from social media. In: *Proceedings of the*

- 6th International AAAI Conference on Weblogs and Social Media, 2012. 483–486
- 28 Zhao Y Y, Qin B, Liu T. Encoding syntactic representations with a neural network for sentiment collocation extraction. *Sci China Inf Sci*, 2017, 60: 110101
- 29 Lakkaraju H, Socher R, Manning C D. Aspect specific sentiment analysis using hierarchical deep learning. In: Proceedings of the NIPS, Workshop on Deep Learning and Representation Learning, 2014
- 30 Xue W, Li T. Aspect based sentiment analysis with gated convolutional networks. In: Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, Melbourne, 2018. 2514–2523
- 31 Taboada M, Brooke J, Tofiloski M, et al. Lexicon-based methods for sentiment analysis. *Comput Linguist*, 2011, 37: 267–307
- 32 Akter S, Aziz M T. Sentiment analysis on facebook group using lexicon based approach. In: Proceedings of International Conference on Electrical Engineering and Information Communication Technology, Bangladesh, 2016
- 33 Diamantini C, Mircoli A, Potena D. A negation handling technique for sentiment analysis. In: Proceedings of International Conference on Collaboration Technologies and Systems (CTS), Orlando, 2016. 188–195
- 34 Mukherjee S, Joshi S. Author-specific sentiment aggregation for polarity prediction of reviews. In: Proceedings of Language Resources and Evaluation Conference, Reykjavik, 2014. 3092–3099
- 35 Perikos I, Hatzilygeroudis I. Aspect based sentiment analysis in social media with classifier ensembles. In: Proceedings of IEEE/ACIS International Conference on Computer and Information Science, 2017. 273–278
- 36 Medhat W, Hassan A, Korashy H. Sentiment analysis algorithms and applications: a survey. *Ain Shams Eng J*, 2014, 5: 1093–1113
- 37 Musto C, Semeraro G, Polignano M. A comparison of lexicon-based approaches for sentiment analysis of microblog posts. In: Proceedings of the 8th International Workshop on Information Filtering and Retrieval Co-located with XIII AI\*IA Symposium on Artificial Intelligence, Pisa, 2014. 59–68
- 38 Khan F H, Qamar U, Bashir S. Lexicon based semantic detection of sentiments using expected likelihood estimate smoothed odds ratio. *Artif Intell Rev*, 2017, 48: 113–138
- 39 Aydogan E, Akcayol M A. A comprehensive survey for sentiment analysis tasks using machine learning techniques. In: Proceedings of International Symposium on Innovations in Intelligent Systems and Applications (INISTA), Sinaia, 2016
- 40 Peng W, Park D H. Generate adjective sentiment dictionary for social media sentiment analysis using constrained nonnegative matrix factorization. In: Proceedings of International AAAI Conference on Weblogs and Social Media, Barcelona, 2011. 273–280
- 41 Povoda L. Sentiment analysis based on support vector machine and big data. In: Proceedings of International Conference on Telecommunications and Signal Processing (TSP), Vienna, 2016. 543–545
- 42 Zainuddin N, Selamat A. Sentiment analysis using support vector machine. In: Proceedings of International Conference on Computer, Communications, and Control Technology (I4CT), Langkawi, 2014
- 43 Pannal N U, Nawarathna C P, Jayakody J T K, et al. Supervised learning based approach to aspect based sentiment analysis. In: Proceedings of IEEE International Conference on Computer and Information Technology (CIT), Nadi, 2016. 662–666
- 44 Appel O, Chiclana F, Carter J, et al. A hybrid approach to sentiment analysis. In: Proceedings of IEEE Congress on Evolutionary Computation (CEC), Vancouver, 2016. 4950–4957
- 45 Mukwazvure A, Supreethi K P. A hybrid approach to sentiment analysis of news comments. In: Proceedings of International Conference on Reliability, Infocom Technologies and Optimization (ICRITO), Noida, 2015
- 46 Goel A, Gautam J, Kumar S. Real time sentiment analysis of tweets using naive bayes. In: Proceedings of International Conference on Next Generation Computing Technologies (NGCT), Dehradun, 2016. 257–261
- 47 Zharmagambetov A S, Pak A A. Sentiment analysis of a document using deep learning approach and decision trees. In: Proceedings of International Conference on Electronics Computer and Computation, Kazakhstan, 2015
- 48 Ouyang X, Zhou P, Li C H, et al. Sentiment analysis using convolutional neural network. In: Proceedings of IEEE International Conference on Computer and Information Technology, 2015. 2359–2364
- 49 Kowsari K, Brown D E, Heidarysafa M. HDLTex: hierarchical deep learning for text classification. In: Proceedings of IEEE International Conference on Machine Learning and Applications, Cancun, 2017. 363–371
- 50 Deng L, Yu D. Deep learning: methods and applications. *FNT Signal Process*, 2013, 7: 197–387
- 51 Mikolov T, Karafiat M, Burget L, et al. Recurrent neural network based language model. In: Proceedings of the 11th Annual Conference of the International Speech Communication Association, 2010
- 52 Mikolov T, Sutskever I, Chen K, et al. Distributed representations of words and phrases and their compositionality. In: Proceedings of International Conference on Neural Information Processing Systems, Lake Tahoe, 2013. 3111–3119
- 53 Luebke D, Humphreys G. How GPUs work. *Computer*, 2007, 40: 96–100
- 54 Yu T, Hidey C, Rambow O, et al. Leveraging sparse and dense feature combinations for sentiment classification. 2017. ArXiv:1708.03940
- 55 Giatsoglou M, Vozalis M G, Diamantaras K, et al. Sentiment analysis leveraging emotions and word embeddings. *Expert Syst Appl*, 2017, 69: 214–224
- 56 Harris Z S. Distributional structure. *WORD*, 1954, 10: 146–162
- 57 Bengio Y, Ducharme R, Vincent P, et al. A neural probabilistic language model. *J Mach Learn Res*, 2003, 3: 1137–1155
- 58 Collobert R, Weston J. A unified architecture for natural language processing. In: Proceedings of International Conference on Machine Learning, Helsinki, 2008. 160–167



- 59 Mikolov T, Chen K, Corrado G, et al. Efficient estimation of word representations in vector space. In: Proceedings of International Conference on Learning Representations (ICLR), Scottsdale, 2013
- 60 Pennington J, Socher R, Manning D C. Glove: global vectors for word representation. In: Proceedings of Conference on Empirical Methods in Natural Language Processing (EMNLP), Doha, 2014. 1532–1543
- 61 Joulin A, Grave E, Bojanowski P, et al. Bag of tricks for efficient text classification. In: Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics, 2017. 427–431
- 62 Zou W Y, Socher R, Cer D, et al. Bilingual word embeddings for phrase-based machine translation. In: Proceedings of Conference on Empirical Methods in Natural Language Processing, Seattle, 2013. 1393–1398
- 63 Devlin J, Chang M W, Lee K, et al. BERT: pre-training of deep bidirectional transformers for language understanding. In: Proceedings of Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Minneapolis, 2019. 4171–4186
- 64 Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need. In: Proceedings of the 31st Annual Conference on Neural Information Processing Systems, Long Beach, 2017. 6000–6010
- 65 Tang D Y, Wei F R, Yang N, et al. Learning sentiment-specific word embedding for twitter sentiment classification. In: Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics, Baltimore, 2014. 1555–1565
- 66 Zhou H W, Chen L, Shi F L, et al. Learning bilingual sentiment word embeddings for cross-language sentiment classification. In: Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing of the Asian Federation of Natural Language Processing, Beijing, 2015. 430–440
- 67 Fu P, Lin Z, Yuan F C, et al. Learning sentiment-specific word embedding via global sentiment representation. In: Proceedings of the 32nd AAAI Conference on Artificial Intelligence, New Orleans, 2018. 4808–4815
- 68 Erhan D, Courville A, Vincent P. Why does unsupervised pre-training help deep learning? *J Mach Learn Res*, 2010, 11: 625–660
- 69 Patterson J, Gibson A. *Deep Learning*. Sebastopol: O'Reilly Media, 2017
- 70 Ballard D H. Modular learning in neural networks. In: Proceeding the 6th National Conference on Artificial Intelligence, Seattle, 1987. 279–284
- 71 Bengio Y. Learning deep architectures for AI. *FNT Mach Learn*, 2009, 2: 1–127
- 72 Vincent P, Larochelle H, Bengio Y, et al. Extracting and composing robust features with denoising autoencoders. In: Proceedings of the 25th International Conference on Machine Learning, Helsinki, 2008. 1096–1103
- 73 Rong W, Nie Y, Ouyang Y, et al. Auto-encoder based bagging architecture for sentiment analysis. *J Visual Lang Comput*, 2014, 25: 840–849
- 74 Zhou H W, Chen L, Huang D G. Cross-lingual sentiment classification based on denoising autoencoder. In: Proceedings of the 3rd CCF Conference on Natural Language Processing and Chinese Computing, Shenzhen, 2014. 181–192
- 75 Hinton G E, Osindero S, Teh Y W. A fast learning algorithm for deep belief nets. *Neur Comput*, 2006, 18: 1527–1554
- 76 Jin Y, Zhang H, Du D L. Improving deep belief networks via delta rule for sentiment classification. In: Proceedings of the 28th International Conference on Tools with Artificial Intelligence, San Jose, 2016. 410–414
- 77 Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative adversarial nets. In: Proceedings of the 27th International Conference on Neural Information Processing Systems, Montréal, 2014. 2672–2680
- 78 Goodfellow I J. Tutorial: generative adversarial networks. In: Proceedings of Neural Information Processing Systems, 2016
- 79 Li Y, Pan Q, Wang S H, et al. A generative model for category text generation. *Inf Sci*, 2018, 450: 301–315
- 80 Vlachostergiou A, Caridakis G, Mylonas P, et al. Learning representations of natural language texts with generative adversarial networks at document, sentence, and aspect level. *Algorithms*, 2018, 11: 164
- 81 Li J, Madry A, Peebles J, et al. On the limitations of first-order approximation in GAN dynamics. In: Proceedings of the 35th International Conference on Machine Learning, Stockholm, 2018. 3011–3019
- 82 Zhang Y Z, Gan Z, Fan K, et al. Adversarial feature matching for text generation. In: Proceedings of the 34th International Conference on Machine Learning, Sydney, 2017. 4006–4015
- 83 Lecun Y, Bottou L, Bengio Y, et al. Gradient-based learning applied to document recognition. *Proc IEEE*, 1998, 86: 2278–2324
- 84 Collobert R, Weston J, Bottou L, et al. Natural language processing (almost) from scratch. *J Mach Learn Res*, 2011, 12: 2493–2537
- 85 Wang J, Sun J Q, Lin H F, et al. Convolutional neural networks for expert recommendation in community question answering. *Sci China Inf Sci*, 2017, 60: 110102
- 86 Kim Y. Convolutional neural networks for sentence classification. In: Proceedings of Conference on Empirical Methods in Natural Language Processing (EMNLP), Doha, 2014. 1746–1751
- 87 Conneau A, Schwenk H, Barrault L, et al. Very deep convolutional networks for text classification. In: Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics, Valencia, 2017. 1107–1116
- 88 Johnson R, Zhang T. Deep pyramid convolutional neural networks for text categorization. In: Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (ACL), Vancouver, 2017. 562–570
- 89 Elman J L. Finding structure in time. *Cogn Sci*, 1990, 14: 179–211
- 90 Ko W J, Tseng B H, Lee H Y. Recurrent neural network based language modeling with controllable external memory.



- In: Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), New Orleans, 2017. 5705–5709
- 91 Hochreiter S. The vanishing gradient problem during learning recurrent neural nets and problem solutions. *Int J Unc Fuzz Knowl Based Syst*, 1998, 06: 107–116
- 92 Hochreiter S, Schmidhuber J. Long short-term memory. *Neural Comput*, 1997, 9: 1735–1780
- 93 Cho K, Merriënboer B V, Gulcehre C, et al. Learning phrase representations using RNN encoder-decoder for statistical machine translation. In: Proceedings of Conference on Empirical Methods in Natural Language Processing (EMNLP), Doha, 2014. 1724–1734
- 94 Bahdanau D, Cho K, Bengio Y. Neural machine translation by jointly learning to align and translate. In: Proceedings of International Conference on Learning Representations (ICLR), San Diego, 2015
- 95 Mishra A, Tamilselvan S, Dasgupta R, et al. Cognition-cognizant sentiment analysis with multitask subjectivity summarization based on annotators' gaze behavior. In: Proceedings of the 32nd AAAI Conference on Artificial Intelligence, New Orleans, 2018, 5884–5891
- 96 Long Y F, Lu Q, Xiang R, et al. A cognition based attention model for sentiment analysis. In: Proceedings of Conference on Empirical Methods in Natural Language Processing, Copenhagen, 2017. 462–471
- 97 Chen H M, Sun M S, Tu C C, et al. Neural sentiment classification with user and product attention. In: Proceedings of Conference on Empirical Methods in Natural Language Processing, Austin, 2016. 1650–1659
- 98 Wang W, Pan S J, Dahlmeier D, et al. Coupled multi-layer attentions for co-extraction of aspect and opinion terms. In: Proceedings of the 31st AAAI Conference on Artificial Intelligence, San Francisco, 2017
- 99 Giannakopoulos A, Musat C, Hossmann A. Unsupervised aspect term extraction with B-LSTM & CRF using automatically labelled datasets. In: Proceedings of the 8th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis, Copenhagen, 2017. 180–188
- 100 Ghosaly D, Akhtary M S, Chauhany D, et al. Contextual inter-modal attention for multi-modal sentiment analysis. In: Proceedings of Conference on Empirical Methods in Natural Language Processing, Brussels, 2018. 3454–3466
- 101 Goller C, Küchler A. Learning task dependent distributed representations by backpropagation through structure. *IEEE Trans Neur Netw*, 1996, 1: 347–352
- 102 Socher R, Pennington J, Huang E H, et al. Semi-supervised recursive autoencoders for predicting sentiment distributions. In: Proceedings of Conference on EMNLP, Scotland, 2011. 151–161
- 103 Socher R, Perelygin A, Wu J, et al. Recursive deep models for semantic compositionality over a sentiment treebank. In: Proceedings of Conference on Empirical Methods in Natural Language Processing, Seattle, 2013. 1631–1642
- 104 Socher R, Manning D C, Ng Y. A. Learning continuous phrase representations and syntactic parsing with recursive neural networks. In: Proceedings of Neural Information Processing Systems, Deep Learning and Unsupervised Feature Learning Workshop, 2010
- 105 Socher R, Manning D C, Huval B, et al. Semantic compositionality through recursive matrix-vector spaces. In: Proceedings of Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning, Jeju Island, 2012. 1201–1211
- 106 Mnih V, Kavukcuoglu K, Silver D, et al. Playing atari with deep reinforcement learning. 2013. ArXiv:abs/1312.5602
- 107 Li Y. Deep reinforcement learning: an overview. 2017. ArXiv:1701.07274
- 108 Mousavi S S, Schukat M, Howley E. Deep reinforcement learning: an overview. In: Proceedings of SAI Intelligent Systems Conference, 2018. 426–440
- 109 Sutton R S, Barto A G. Reinforcement Learning: an Introduction. Cambridge: MIT Press, 2018
- 110 Mnih V, Kavukcuoglu K, Silver D, et al. Human-level control through deep reinforcement learning. *Nature*, 2015, 518: 529–533
- 111 Williams R J. Simple statistical gradient-following algorithms for connectionist reinforcement learning. *Mach Learn*, 1992, 8: 229–256
- 112 Zhang T Y, Huang M L, Zhao L. Learning structured representation for text classification via reinforcement learning. In: Proceedings of AAAI Conference on Artificial Intelligence, New Orleans, 2018
- 113 Yogatama D, Blunsom P, Dyer C, et al. Learning to compose words into sentences with reinforcement learning. In: International Conference on Learning Representations (ICLR), Toulon, 2017
- 114 Chen M, Wang S, Liang P P, et al. Multimodal sentiment analysis with word-level fusion and reinforcement learning. In: Proceedings of the 19th ACM International Conference on Multimodal Interaction (ICMI), Glasgow, 2017. 163–171
- 115 Su P-H, Gašić M, Mrkšić N, et al. On-line active reward learning for policy optimisation in spoken dialogue systems. In: Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, Berlin, 2016. 2431–2441
- 116 Kamthe S, Deisenroth M P. Data-efficient reinforcement learning with probabilistic model predictive control. In: Proceedings of the 21st International Conference on Artificial Intelligence and Statistics, Lanzarote, 2018. 1701–1710
- 117 Feng X C, Qin B, Liu T. A language-independent neural network for event detection. *Sci China Inf Sci*, 2018, 61: 092106
- 118 Gonzlez J, Pla F, Hurtado L. ELiRF-UPV at SemEval-2017 task 4: sentiment analysis using deep learning. In: Proceedings of the 11th International Workshop on Semantic Evaluations (SemEval), Vancouver, 2017. 723–727
- 119 Tang D Y, Qin B, Liu T. Document modeling with gated recurrent neural network for sentiment classification. In: Proceedings of Conference on Empirical Methods in Natural Language Processing, Lisbon, 2015. 1422–1432

- 120 Maas A L, Daly R E, Pham P T, et al. Learning word vectors for sentiment analysis. In: Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, Portland, 2011. 142–150
- 121 Diao Q, Qiu M, Wu C-Y. Jointly modeling aspects, ratings and sentiments for movie recommendation (JMARS). In: Proceedings of the 20th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, New York, 2014. 193–202
- 122 McAuley J, Leskovec J. Addressing complex and subjective product-related queries with customer reviews. In: Proceedings of the 25th International World Wide Web Conference, Montréal, 2016. 625–635
- 123 Li X, Bing L D, Lam W, et al. Transformation networks for target-oriented sentiment classification. In: Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, Melbourne, 2018. 946–956
- 124 Go A, Bhayani R, Huang L. Twitter sentiment classification using distant supervision. 2009. <http://pdfs.semanticscholar.org/52e2/bd533323ddf97073d034bae40a46eda55f34.pdf>
- 125 Saif H, Fernandez M, He Y, et al. Evaluation datasets for Twitter sentiment analysis: a survey and a new dataset, the STS-Gold. In: Proceedings of the 1st International Workshop on Emotion and Sentiment in Social and Expressive Media: Approaches and Perspectives from AI, Turin, 2013. 9–21
- 126 Zadeh A, Zellers R, Pincus E, et al. MOSI: multimodal corpus of sentiment intensity and subjectivity analysis in online opinion videos. 2016. ArXiv:1606.06259
- 127 Zadeh A, Liang P P, Vanbriesen J, et al. Multimodal language analysis in the wild: CMU-MOSEI dataset and interpretable dynamic fusion graph. In: Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, Melbourne, 2018. 2236–2246
- 128 Pang B, Lee L. Seeing stars: exploiting class relationships for sentiment categorization with respect to rating scales. In: Proceedings of the Association for Computational Linguistics, 2005. 115–124
- 129 Sanders N J. Sanders-Twitter sentiment corpus. Sanders Analytics LLC, 2011. [http://www.scirp.org/\(S\(lz5mqp453edsnp55rrgjt55\)\)/reference/ReferencesPapers.aspx?ReferenceID=1613287](http://www.scirp.org/(S(lz5mqp453edsnp55rrgjt55))/reference/ReferencesPapers.aspx?ReferenceID=1613287)
- 130 Wojatzki M, Ruppert E, Holschneider S, et al. GermEval 2017: shared task on aspect-based sentiment in social media customer feedback. In: Proceedings of the GermEval 2017 Shared Task, Berlin, 2017
- 131 Nabil M, Aly M, Atiya A F. ASTD: arabic sentiment tweets dataset. In: Proceedings of Conference on Empirical Methods in Natural Language Processing, Lisbon, 2015. 2515–2519
- 132 Yang Z C, Yang D Y, Dyer C. Hierarchical attention networks for document classification. In: Proceedings of Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, San Diego, 2016. 1480–1489
- 133 Ganin Y, Ustinova E, Ajakan H, et al. Domain-adversarial training of neural networks. *J Mach Learn Res*, 2016, 27: 1–35
- 134 Li Z, Zhang Y, Wei Y, et al. End-to-end adversarial memory network for cross-domain sentiment classification. In: Proceedings of the 26th International Joint Conference on Artificial Intelligence (IJCAI), Melbourne, 2017. 2237–2243
- 135 Zhai S F, Zhang Z F. Semisupervised autoencoder for sentiment analysis. In: Proceedings of the 30th AAAI Conference on Artificial Intelligence, Phoenix, 2016. 1394–1400
- 136 Yin R C, Li P, Wang B. Sentiment lexical-augmented convolutional neural networks for sentiment analysis. In: Proceedings of the 2nd International Conference on Data Science in Cyberspace (DSC), Shenzhen, 2017. 630–635
- 137 Tai K S, Socher R, Manning C D. Improved semantic representations from tree-structured long short-term memory networks. In: Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing, Beijing, 2015. 1556–1566
- 138 Kokkinos F, Potamianos A. Structural attention neural networks for improved sentiment analysis. In: Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics, Valencia, 2017. 586–591
- 139 Mousa A, Schuller B. Contextual bidirectional long short-term memory recurrent neural network language models: a generative approach to sentiment analysis. In: Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics, Valencia, 2017. 1023–1032
- 140 Wang Y Q, Huang M L, Zhao L, et al. Attention-based LSTM for aspect-level sentiment classification. In: Proceedings of Conference on Empirical Methods in Natural Language Processing, Austin, 2016. 606–615
- 141 Tay Y, Tuan L A, Hui S C. Learning to attend via word-aspect associative fusion for aspect-based sentiment analysis. In: Proceedings of the 32nd AAAI Conference on Artificial Intelligence, New Orleans, 2018. 5956–5963
- 142 Zhang M S, Zhang Y, Vo D. Gated neural networks for targeted sentiment analysis. In: Proceedings of the 30th AAAI Conference on Artificial Intelligence, Phoenix, 2016. 3087–3093
- 143 Yang J, Yang R Q, Wang C J, et al. Multi-entity aspect-based sentiment analysis with context, entity and aspect memory. In: Proceedings of the 32nd AAAI Conference on Artificial Intelligence, New Orleans, 2018. 6029–6036
- 144 He R, Lee W S, Ng H T, et al. Effective attention modeling for aspect-level sentiment classification. In: Proceedings of the 27th International Conference on Computational Linguistics, Santa Fe, 2018. 1121–1131
- 145 Chen P, Sun Z Q, Bing L D, et al. Recurrent attention network on memory for aspect sentiment analysis. In: Proceedings of Conference on Empirical Methods in Natural Language Processing (EMNLP), Copenhagen, 2017. 452–461
- 146 Ma Y K, Peng H Y, Cambria E. Targeted aspect-based sentiment analysis via embedding commonsense knowledge into an attentive LSTM. In: Proceedings of the 32nd AAAI Conference on Artificial Intelligence, New Orleans, 2018.

5876–5883

- 147 Huang B, Carley K M. Parameterized convolutional neural networks for aspect level sentiment classification. In: Proceedings of Conference on Empirical Methods in Natural Language Processing, Brussels, 2018. 1091–1096
- 148 Jakob N, Gurevych I. Extracting opinion targets in a single- and cross-domain setting with conditional random fields. In: Proceedings of Empirical Methods in Natural Language Processing, Massachusetts, 2010. 1035–1045
- 149 Wang W, Pan S J, Dahlmeier D, et al. Recursive neural conditional random fields for aspect-based sentiment analysis. In: Proceedings of Conference on Empirical Methods in Natural Language Processing, Austin, 2016. 616–626
- 150 Guo L D, Jiang S Y, Du W J, et al. Recurrent neural CRF for aspect term extraction with dependency transmission. In: Proceedings of CCF International Conference on Natural Language Processing and Chinese Computing, 2018. 378–390
- 151 Xue W, Zhou W B, Li T, et al. MTNA: a neural multi-task model for aspect category classification and aspect term extraction on restaurant reviews. In: Proceedings of the 8th International Joint Conference on Natural Language Processing, Taipei, 2017. 151–156
- 152 Li X, Lam W. Deep multi-task learning for aspect term extraction with memory interaction. In: Proceedings of Conference on Empirical Methods in Natural Language Processing, Copenhagen, 2017. 2886–2892
- 153 Li X, Bing L D, Li P J. Aspect term extraction with history attention and selective transformation. In: Proceedings of the 27th International Joint Conference on Artificial Intelligence, Stockholm, 2018. 4194–4200
- 154 Ren Y F, Ji D H. Neural networks for deceptive opinion spam detection: an empirical study. *Inf Sci*, 2017, 385: 213–224
- 155 Bandhakavi A, Wiratunga N, Padmanabhan D, et al. Lexicon based feature extraction for emotion text classification. *Pattern Recogn Lett*, 2017, 93: 133–142
- 156 Felbo B, Mislove A, Søgaard A. Using millions of emoji occurrences to learn any-domain representations for detecting sentiment, emotion and sarcasm. In: Proceedings of Conference on Empirical Methods in Natural Language Processing, Copenhagen, 2017. 1615–1625
- 157 Kim Y, Lee H. AttnConvnet at SemEval-2018 task 1: attention-based convolutional neural networks for multi-label emotion classification. In: Proceedings of the 12th International Workshop on Semantic Evaluation, New Orleans, 2018. 141–145
- 158 Yu J F, Marujo L, Jiang J, et al. Improving multi-label emotion classification via sentiment classification with dual attention transfer network. In: Proceedings of Conference on Empirical Methods in Natural Language Processing, Brussels, 2018. 1097–1102
- 159 Wang Y Q, Feng S, Wang D L, et al. Multi-label chinese microblog emotion classification via convolutional neural network. In: Proceedings of Asia-Pacific Web Conference, 2016. 567–580
- 160 Abdul-Mageed M, Ungar L. EmoNet: fine-grained emotion detection with gated recurrent neural networks. In: Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics, Vancouver, 2017. 718–728
- 161 Tafreshi S, Diab M. Emotion detection and classification in a multigenre corpus with joint multi-task deep learning. In: Proceedings of the 27th International Conference on Computational Linguistics, Santa Fe, 2018. 2905–2913
- 162 Chen X L, Sun Y, Athiwaratkun B, et al. Adversarial deep averaging networks for cross-lingual sentiment classification. *Trans Assoc Comput Linguist*, 2018, 6: 557–570
- 163 Ruder S, Ghaffari P, Breslin J G. INSIGHT-1 at SemEval-2016 task 5: deep learning for multilingual aspect-based sentiment analysis. In: Proceedings of SemEval, San Diego, 2016. 330–336
- 164 Becker W, Wehrmann J, Cagnini H E L, et al. An efficient deep neural architecture for multilingual sentiment analysis in Twitter. In: Proceedings of the 30th International Florida Artificial Intelligence Research Society Conference, Florida, 2017. 246–251
- 165 Attia M, Samih Y, ElKahky A, et al. Multilingual multi-class sentiment classification using convolutional neural networks. In: Proceedings of the 11th International Conference on Language Resources and Evaluation (LREC), Miyazaki, 2018
- 166 Can E F, Ezen-Can A, Can F. Multilingual 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, 2018
- 167 Akhtar M S, Sawant P, Sen S, et al. Solving data sparsity for aspect based sentiment analysis using cross-linguality and multi-linguality. In: Proceedings of NAACL-HLT, New Orleans, 2018. 572–582
- 168 Wang W C, Feng S, Gao W. Personalized microblog sentiment classification via adversarial cross-lingual multi-task learning. In: Proceedings of Conference on Empirical Methods in Natural Language Processing, Brussels, 2018. 338–348
- 169 Cambria E, Hazarika D, Poria S, et al. Benchmarking multimodal sentiment analysis. In: Proceedings of International Conference on Computational Linguistics, Budapest, 2017. 166–179
- 170 Poria S, Cambria E, Hazarika D. Context-dependent sentiment analysis in user-generated videos. In: Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics, Vancouver, 2017. 873–883
- 171 Zadeh A, Chen M, Poria S. Tensor fusion network for multimodal sentiment analysis. In: Proceedings of Conference on Empirical Methods in Natural Language Processing, Copenhagen, 2017. 1103–1114
- 172 Zadeh A, Liang P P, Poria S, et al. Multi-attention recurrent network for human communication comprehension. In: Proceedings of the 32nd AAAI Conference on Artificial Intelligence, New Orleans, 2018. 5642–5649
- 173 Wu H B, Gu Y W, Sun S D, et al. Aspect-based opinion summarization with convolutional neural networks. In: Proceedings of International Joint Conference on Neural Networks, 2016. 3157–3163

- 174 Li Q D, Jin Z P, Wang C, et al. Mining opinion summarizations using convolutional neural networks in Chinese microblogging systems. *Knowl-Based Syst*, 2016, 107: 289–300
- 175 Wang L, Ling W. Neural network-based abstract generation for opinions and arguments. In: *Proceedings of NAACL-HLT*, San Diego, 2016. 47–57
- 176 Yang M, Qu Q, Zhu J, et al. Cross-domain aspect/sentiment-aware abstractive review summarization. In: *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*, Torino, 2018. 1531–1534
- 177 Yang M, Qu Q, Shen Y. Aspect and sentiment aware abstractive review summarization. In: *Proceedings of the 27th International Conference on Computational Linguistics*, Santa Fe, 2018. 1110–1120
- 178 Angelidis S, Lapata M. Summarizing opinions: aspect extraction meets sentiment prediction and they are both weakly supervised. In: *Proceedings of Conference on Empirical Methods in Natural Language Processing*, Brussels, 2018. 3675–3686
- 179 Zhao S Y, Xu Z W, Liu L M, et al. Towards accurate deceptive opinions detection based on word order-preserving CNN. *Math Probl Eng*, 2018, 2018: 1–9
- 180 Li L Y, Qin B, Ren W J, et al. Document representation and feature combination for deceptive spam review detection. *Neurocomputing*, 2017, 254: 33–41
- 181 Ren Y F, Yue Z. Deceptive opinion spam detection using neural network. In: *Proceedings of the 26th International Conference on Computational Linguistics*, Osaka, 2016. 140–150
- 182 Zhang W, Du Y H, Yoshida T, et al. DRI-RCNN: an approach to deceptive review identification using recurrent convolutional neural network. *Inf Process Manage*, 2018, 54: 576–592
- 183 Yadav V, Bethard S. A survey on recent advances in named entity recognition from deep learning models. In: *Proceedings of the 27th International Conference on Computational Linguistics*, Santa Fe, 2018. 2145–2158
- 184 Qiu J Z, Tang J, Ma H, et al. DeepInf: social influence prediction with deep learning. In: *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, London, 2018. 2110–2119
- 185 Wang H W, Zhang F Z, Hou M, et al. SHINE: signed heterogeneous information network embedding for sentiment link prediction. In: *Proceedings of the 11th ACM International Conference on Web Search and Data Mining*, Marina Del Rey, 2018. 592–600