Computer Science Review 49 (2023) 100576

Contents lists available at ScienceDirect

Computer Science Review

journal homepage: www.elsevier.com/locate/cosrev

Review article

Aspect based sentiment analysis using deep learning approaches: A survey

Ganpat Singh Chauhan [a,*], Ravi Nahta [b], Yogesh Kumar Meena [c], Dinesh Gopalani [c]

[a] *Department of Information Technology, School of IT, Manipal University Jaipur, India*
[b] *Department of Computer Science & Engineering, Nirma University, India*
[c] *Department of Computer Science & Engineering, Malaviya National Institute of Technology, Jaipur, Rajasthan, India*

ARTICLE INFO

Article history:
Received 7 January 2022
Received in revised form 7 July 2023
Accepted 25 July 2023
Available online 6 August 2023

Keywords:
Explicit aspect extraction
Aspect category detection
Attention
Deep neural network

ABSTRACT

The wealth of unstructured text on the online web portal has made opinion mining the most thrust area for researchers, academicians, and businesses to extract information for gathering, analyzing, and aggregating human emotions. The extraction of public sentiment from the text at an aspect level has contributed exceptionally to various businesses in the marketplace. In recent times, deep learning-based techniques have learned high-level linguistic features without high-level feature engineering. Therefore, this paper focuses on a rigorous survey on two primary subtasks, aspect extraction and aspect category detection of aspect-based sentiment analysis (ABSA) methods based on deep learning. The significant advancement in the ABSA sector is demonstrated by a thorough evaluation of state-of-the-art and latest aspect extraction methodologies.

Contents

* Corresponding author.
E-mail addresses: ganpatchauhan@gmail.com (G.S. Chauhan), ravi.nahta@nirmauni.ac.in (R. Nahta), ymeena.cse@mnit.ac.in (Y.K. Meena), dgopalani.cse@mnit.ac.in (D. Gopalani).

https://doi.org/10.1016/j.cosrev.2023.100576
1574-0137/

1. Introduction

In the modern digital era of information society, the vast amount of user content generated from ever-increasing internet and online activities has opened a new direction for communication through unstructured text (chatting, blogs, reviews, social media content, e-commerce, online banking, etc.). If processing is carried out manually, managing this enormous content takes a lot of time. One of the most popular methods for extracting sentiment from unstructured text is sentiment analysis. A sentimental analysis of social networking websites was done to reduce the traditional methods being used for gathering suggestions and comments. The majority of the processing of user comments has just involved putting the tone into negative or positive categories. Moreover, the information shared by the users on the online portal is more trustworthy than the information provided by the vendors. Thus, knowing the likes and dislikes of potential users could contribute significantly to maintaining product quality, improving existing ones, developing new products, and judging society's impact. Hu and Liu [1] argued that there are three distinct classifications of sentiment analysis (SA): (1) document-level, (2) sentence level as given in Fig. 1, and (3) aspect-level [2].

In recent research, the online material has been acknowledged as a largest source of information for understanding the sentiments of the general public for a variety of application domains like products, restaurants, movies, political issues, educational plans [3], social events, YouTube video ranking [4], market campaigns, and government policies. Using a multi-channel convolutional network, [5] execute ABSA on microblogs and news articles in the economic area [6,7].

The remainder of the article is structured in the following manner: Section 2 discusses state-of-the-art machine learning-based and lexicon-based sentiment classification approaches. Section 3 defines subtasks of the ABSA, standard datasets used, and performance measures for the ABSA. Next, in Section 4, we discuss conventional aspect-based sentiment analysis approaches. Further, Section 5 explains the deep learning approach for ABSA. Furthermore, Section 6 describes the input vectors such as CBOW, skipgrams, and Bidirectional Encoder Representations from Transformers (BERT) used as embedding layers in deep neural networks for aspect term extraction. In order to extract the aspect subtask in ABSA, Section 7 derives the operation of the models used for deep learning, particularly the Convolutional Neural Network (CNN), the Recurrent Neural Network (RNN), and the attention technique. Section 8 presents the performance analysis and impact of various deep learning models and hybrid models for various problem areas such as domain adaptation issues, aspect extraction from multi-sentence reviews, etc. In addition, this section also presents the impact of attention mechanisms and BERT. Section 9 outlines the principal difficulties in deep learning-based aspect extraction as well as ABSA's future direction. Finally, the survey on deep learning approaches for ABSA is summarized in the conclusion presented in Section 10.

2. Sentiment classification approaches

The opinion classification approaches shown in Fig. 1 [8] are primarily divided into lexicon-based techniques and machine learning approaches. Machine Learning (ML) applies semantic features, whereas the lexicon technique depends on an opinion lexicon (an opinion term collection). In sentiment classification methods, both supervised and unsupervised learning are used in machine learning systems. An extensive collection of labeled training review phrases is assembled by the supervised approaches. In contrast, *unsupervised methods* are applied if the labeled training reviews are not available. The dictionary-based approach identifies opinion seed phrases and looks up synonyms and antonyms in the lexicon. The corpus-based approach searches a large corpus for additional sentiment words and extracts sentiment words with domain-related orientations using a seed list of opinion terms. Finally, the corpus-based lexicon approach applies semantic or statistical methods to extract opinion polarity.

2.1. Machine learning approaches

The supervised learning methods require labeled training reviews. In the literature, probabilistic classifiers, linear classifiers, decision trees, and rule-based classifiers are the most often employed supervised classifiers for sentiment analysis [8]. The *probabilistic classifiers* apply combined models for sentiment extraction. This combined model presumes each class as a combination's element. Here, the individual element is a generative model that imparts a particular term's sampling probability to a specific element. Probabilistic classifiers are also known as generative classifiers. The most popular probabilistic classifier is the Naive Bayes (*NB*) classifier. It calculates the class posterior probability on the basis of the word distribution in the review sentence. *NB* applies the Bayes Theorem to forecast the opinion probability of the sentence review. The major challenge that the *NB* classifier should expect is the feature's autonomy. Another main expectation of NB is to assume fully dependent features, and the *Bayesian Network* (*BN*) model exactly does the same. However, the computation complexity of *BN* is extremely high; therefore, the researchers are not commonly using it. The *Maximum Entropy* (*ME*) Classifier uses encoding to transform labeled feature sets into vectors.

In *Linear classifiers*, the output is represented as the linear predictor output, $p = A'.X'$; where X' is the sentencing word frequency, A' represents linear coefficient vector, whereas b denotes scalar. Here, p separates the hyperplanes between two sentiment classes. The most popular linear classifiers are support vector machines (*SVM*) and neural networks (*NN*). The major concept of *SVMs* is to find out the search space's linear separators that optimally distinguish various classes. The sparse behavior of textual data is preferably complemented by *SVM* classification. Here, a few irrelevant features match up altogether and are mostly organized into linearly separable classes. *SVM* establishes a non-linear decision plane in the initial feature space by non-linearly translating the input groups. The linear neural network function is: p_i = A.X_i'; where X_i' is the word frequency in the ith review

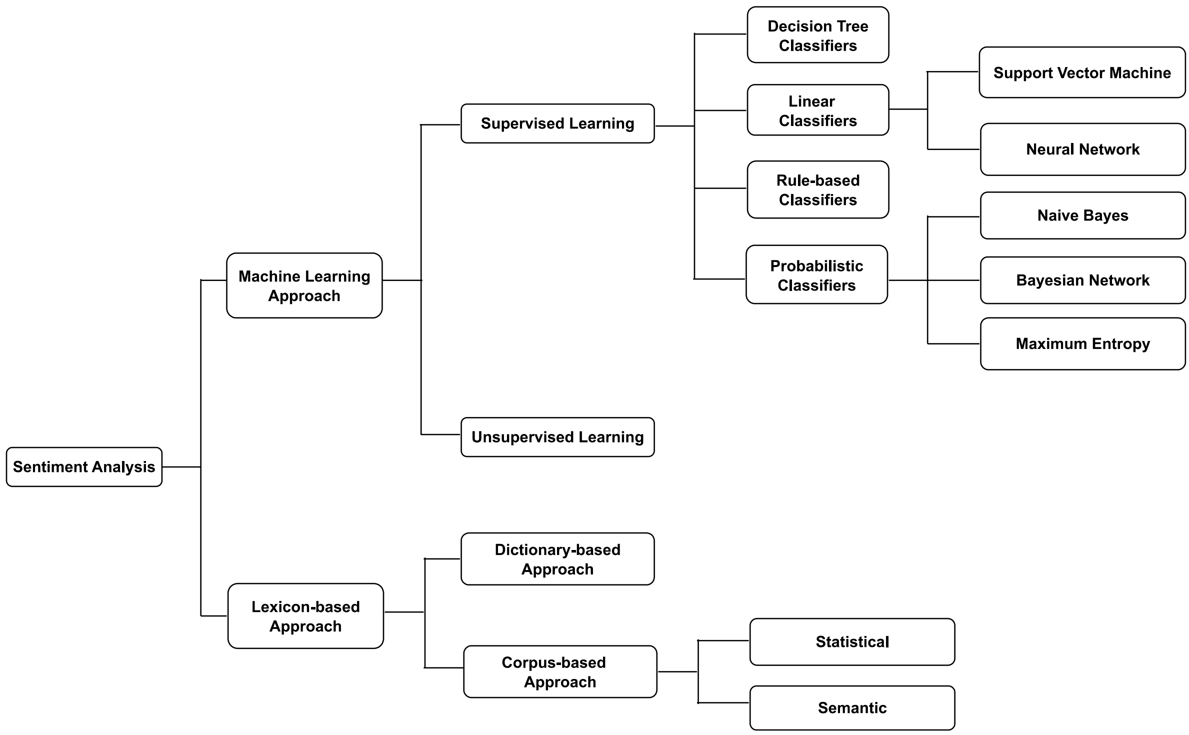

Fig. 1. Techniques for classifying sentiments taken from [8].

sentence and the class label for X_i is denoted as y_i, A is the weight set for each neuron p_i is for the label class. The non-linear distinction between classes is accomplished using multilayer *NN*. These several layers produce a variety of piecewise linear borders that roughly represent enclosed regions which belong to a particular class. Neurons in later layers get input from neurons in previous layers via their outputs. Because errors must be back-propagated through multiple layers, the training process is more difficult [8]. According to the literature, *SVM* has been widely used for opinion mining, whereas NNs have received minimal attention as a classifier for opinion analysis.

The *decision tree* (*DT*) classifier comes up with a hierarchical breakdown of the data used for training space, where the prediction value of the feature divides the data. The predicate represents the existence or non-existence of nouns or noun phrases. The dataset separation works iteratively until the leaf nodes hold precise records for the classification task. *Rule-based* classifiers use various rules for data space modeling. The right component of the feature set represents the class label, while the left part serves as a prediction. The training phase establishes the regulations based on *support* and *confidence* standards. The *support* denotes absolute instance numbers that are relevant to the rule in the training sample. The *confidence* indicates the conditional likelihood which the right part of the rule will be fulfilled when the left component is satisfied. The primary distinction between decision rules and decision trees is that while rule-based classifiers permit overlapping in the context of decision space, *DT* strictly partitions the data space hierarchically [9,10].

2.2. Lexicon-based approaches

Generally, the sentiment classification subtask uses opinion words. Positive opinion words express a few expected states, whereas negative sentiment words express some unexpected states. The combined sentiment phrases and idioms are known as the *sentiment lexicon*. A manual approach to compiling or capturing the sentiment word list is highly recommended. Therefore, it is generally combined with automated approaches like *dictionary-based* or *corpus-based* for final verification to keep away the errors resulting from the automated methods. The dictionary-based method involves manually compiling a small set of emotive phrases with predetermined orientations. The widely recognized corpus WordNet, or thesaurus is then used to augment this collection by providing synonyms and antonyms. They also discovered more words, put them on the seed list, and the iteration process started. Whenever no novel terms are discovered, the iterative procedure eventually comes to an end. Once the procedure is complete, errors can be found or precisely identified via manual inspection. However, the dictionary-based approach cannot find domain-specific sentiment word orientation. The corpus-based approach overcomes the issue of domain-specific sentiment word orientation. Its methods rely on syntactic patterns to acquire more opinion phrases within a huge corpus. Applying only the corpus-driven strategy is not as effective compared to the dictionary-based technique. The reason is that creating a huge corpus that covers all English words is very difficult. However, this method excels at extracting sentiment words that are exclusive to a given domain. The corpus-based technique uses a statistical approach or semantic methods for sentiment extraction. Extracting repeated patterns or seed sentiment words is performed with statistical methods.

The simultaneous occurrences of corpus adjectives are used for pattern extraction by obtaining posterior polarities. Thus, it is feasible to utilize the entire collection of indexed web review sentences as the repository for the creation of a dictionary. This approach reduces few-word deficiency issues for small corpora. The frequency of occurrence of the word in the extensive annotated corpora of documents can be used to identify the word's polarity. If the word is frequently accompanied by a high number of positive reviews, then the polarity is positive. Conversely, if the word is more frequent in negative reviews, then the polarity is negative. In the case of equal frequencies, the word has

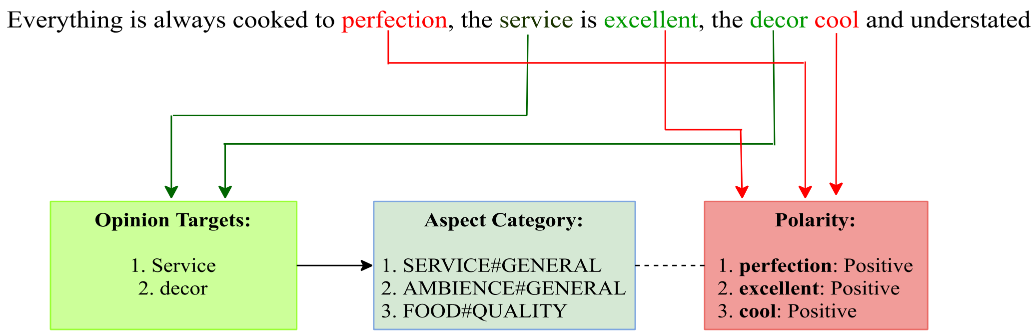

Fig. 2. An example for three subtasks of ABSA.

neutral polarity. The same sentimental words are often placed together in a corpus. Therefore, if two words frequently occur together inside an identical domain, the polarity must be the same. The semantic methodology directly provides opinion polarities, which depend upon various concepts in evaluating the word's similarity. These concepts give identical opinion values to conceptually related words. For example, WordNet imparts various kinds of conceptual relationships among opinion polarities. In addition, the WordNet iteratively enlarges the first set of terms with synonyms and antonyms to create a list of words with opinions. These words were further utilized concerning the count of positive and negative synonyms while establishing a word's emotional spectrum for the first time.

The coarse-grained sentiment analysis approaches discussed above deliver outstanding outcomes across several application disciplines. Even so, a general polarity may not always accurately capture the text's meaning. Therefore, it is essential to capture higher-level information and contextual knowledge to extract various aspect terms. Moreover, with the rapid growth of text, organizations and individuals are interested in finding the opinion's root cause. The user comments and reviews contain views about different aspects discussed within the review phrase.

3. Aspect based sentiment analysis (ABSA)

From the last decade, a lot of literature has been published on various granularity levels of sentiment analysis. The sentiment extraction may be defined formally as the quadruple (p, e, o, t) finding, where p denotes the opinion polarity, e means the target entity, o stands for the sentiment bearer, and t represents the sentiment conveyed time. The SA methods emphasize mostly on extracting the pair (p, e). The intended object e could be an entity (overall topic) or an aspect (attribute of the topic discussed) [11]. The document-level or sentence-level sentiment analysis focuses only on one subject of the review [2,12,13] have presented a rigorous survey on elementary knowledge, different applications, and approaches for sentiment analysis problems covering practical and ethical scenarios [14]. These surveys focus on machine learning, dictionary-based, statistical, and semantic strategies for recognizing the text's emotion at the phrase and document levels. A theoretical analysis of different concepts and techniques in sentiment analysis and subjectivity has been presented [15]. An overall polarity sometimes does not reflect the exact sentiment of the text. With the rapid growth of text, organizations and individuals are interested in finding the opinion's root cause.

3.1. Subtasks of ABSA

As given in Fig. 2, the three main subtasks of ABSA are as follows: Extracting aspect-opinion pairs (ATE); identifying aspect categories (ACD), which classifies the aspect-opinion pair; and (iii) sentiment polarity (sentiment aggregation) [1,16,17]. The first step (ATE) extracts aspect terms (entities and attributes) and their associated opinion terms from the text. The next step (ACD) is to categorize the retrieved aspects into a predetermined range of aspect categories. Finally, a summarization of the polarity of sentiment words is done for each of the extracted aspects to provide an aggregate overview. Let us consider the review given in Fig. 2 for the restaurant domain to understand the three subtasks of ABSA. *FOOD, SERVICE,* and *AMBIENCE* are entities in this review, whereas *service* and *decor* are aspect terms. *Perfection*, *excellent*, and *cool* belong to sentiment words that target the entities and aspects extracted from the review. For ACD, the task is to identify categories for the extracted opinion targets, which are: *FOOD#QUALITY* (for *FOOD*), *SERVICE#SERVICE* (for *SERVICE*), and AMBIENCE#GENERAL (for *DECOR*). Sentiment polarity extracts the feeling behind each aspect category. Both *service* and *excellent*, as well as *decor* and *cool*, are related to the ABSA tasks of aspect category recognition and sentiment polarity.

3.2. Datasets for ABSA

The techniques used in the taxonomy of aspect extraction are based on a variety of domains and language datasets. The customer reviews dataset from different online platforms like Yelp, Amazon, etc. has been focused on most techniques. These datasets are majorly to be restaurant, service, movie, or product review domains, which consider that each review holds one entity with one or multiple aspects. The benchmark datasets available among research communities for ABSA include customer reviews of digital products [1], as shown in Table 1. SemEval-14 for restaurants and laptops [20], target-dependent Twitter sentiment classification dataset for Twitter comments [21], SemEval-16 as shown in Table 2 for Restaurants, and laptops [16], ICWSM JDPA sentiment corpus for automotive and digital devices, and Darmstadt Service review Corpus for online universities and online service reviews [22]; FIQA ABSA for financial news headlines and financial microblogs [23]. Most of these datasets have been annotated to indicate whether each sentence contains an aspect or not. Each review that contains the aspect term (s) is labeled with aspect, categories of aspect, and polarity.

In recent years, in the era of deep learning-based techniques for ABSA, the SemEval-14 and SemEval-16 datasets released by the 'International Workshop on Semantic Evaluation' have been targeted by researchers and academicians. The statistics of the SemEval-16 dataset are given in Table 3 and Table 4. This survey performed a detailed performance analysis of recent deep learning-based ABSA approaches on restaurant and laptop domains for the SemEval-14 and SemEval-16 datasets.

Table 1
Statistics of Hu and Liu [1] product datasets.

Datasets ->	Canon digital camera	Nikon digital camera	Apex DVD player	Creative MP3 player	Nokia cell phone
# reviews	45	34	41	95	99
# sentences	597	346	546	1716	740
# of opinionated reviews	238	160	265	720	344
# of non-opinionated reviews	359	186	282	996	396
# aspects	237	174	302	674	296

Table 2
Statistics of SemEval-16 dataset [18].

Criteria	Restaurant			Laptop		
	Training data	Testing data	Total	Training data	Testing data	Total
Reviews	395	91	486	451	80	531
Sentences	2000	676	2676	2500	800	3300
Aspect category	12	12	12	81	68	88

Table 3
Statistics of SemEval-16 for some aspect categories (Restaurant, Laptop) [19].

Aspect category	Training	Testing	Criteria (in sentences)	Restaurant	Laptop
AMBIENCE#GENERAL	255	66	# with multiple category	540	688
DRINKS#QUALITY	47	22	# with single category	1189	1377
FOOD#PRICES	90	23	# different categories for same aspects	122	0
FOOD#QUALITY	849	313	# same categories for different aspects	190	0
FOOD#STYLE_OPTIONS	137	55	# without category	271	435
RESTAURANT#GENERAL	422	142	# sentences with background (intra-sentence dependency)	315	454
SERVICE#GENERAL	449	155	# of reviews with Elaboration (inter-sentence dependency)	57	49
LAPTOP#GENERAL	634	158			

Table 4
Aspects statistics of SemEval-16(Restaurant) [24].

Criteria	Training	Testing
Total reviews	395	91
Total sentences	2000	676
# of aspect sentences	1254	414
# of non-aspect sentences	746	264
# of sentence single word aspects	898	331
#. of sentence with nouns phrase	356	83
# of sentences with multiple category	901	347
# of sentence with multiple aspects	353	67
Total # of aspects	2507	651
# of unique aspects	635	265
Total noun phrase aspect	379	135
Total single word aspect	256	130

3.3. Evaluation measures

Precision, recall, and f-score are considered evaluation metrics for varieties of aspect extraction techniques. The Semantic Evaluation Workshops share SemEval datasets [20] and [16] among all the participants to promote aspect-based sentiment analysis as an evaluation methodology. Precision, recall, and F-score are evaluated using true-negative (*TN*), true positive (*TP*), false-positive (*FP*), and false-negative (*FN*) values for the model.

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Positive} \tag{1}$$

$$Recall = \frac{True\ Positive}{True\ Positive + False\ Negative} \tag{2}$$

$$F\text{-}Score = \frac{2 * Precision * Recall}{Precision + Recall} \tag{3}$$

This article has used a set A of extracted aspects where T is a set of manually annotated aspects from the dataset, to calculate the above values. Hence, *TP*, *FP* and *FN* will respectively be $|A \cap T|$, $|A \setminus T|$, and $|T \setminus A|$ [25]. *TP* and *FP* represent the labels, whereas the model assigns non-labels to *TN*. In contrast, some labels are manually annotated, which were missed by the model. Precision denotes the percentage of correctly extracted labels (exactness or quality); whereas recall represents the completeness or quantity of extracted labels. In simple terms, higher *precision* means the system extracts substantially more relevant aspects than irrelevant ones. On the other side, high *recall* means the system extracts most of the relevant aspects. The true results (*TN* and *TF*) represent the F-score and accuracy. In most of the methodologies, the ATE and ACD subtasks have used the F-score to evaluate sentiment aggregation and task accuracy.

4. Conventional aspect-based sentiment analysis approaches

This paper focuses on aspect-level sentiment analysis for explicit aspect extraction subtasks: aspect and opinion extraction, joint aspect detection, and sentiment analysis. As depicted in Fig. 3, there are several different aspect extraction techniques, including deep learning, supervised and unsupervised machine learning, frequency-based, syntax-based, and hybrid techniques [11]. *Frequency-based* methods have noticed that a particular set of words appears considerably more frequently than other vocabulary words in review sentences. These frequently used terms are expected aspects. This simple frequency method has been proven to be quite powerful; still, the drawbacks are that all frequent nouns and noun phrases do not always represent aspects. In addition, frequency-based methods may overlook some details that are not often mentioned in reviews. The frequency-based methods can extend the above issues. Hence, these manually drafted rules often require fine-tuning. Alternatively, *syntax-based* methods extract aspects by using syntactical relations rather than focusing on the frequencies of aspects. As in "tasty rice," where "tasty" is an adjective modifying the aspect "rice," this relationship between a sentimental word and an aspect term is prevalent. The ability to efficiently prune away low-frequency characteristics is a key idea of syntax-based techniques. Nevertheless, various syntactical associations are required to obtain adequate coverage. Hence, this method extracts additional sentiment-aspect

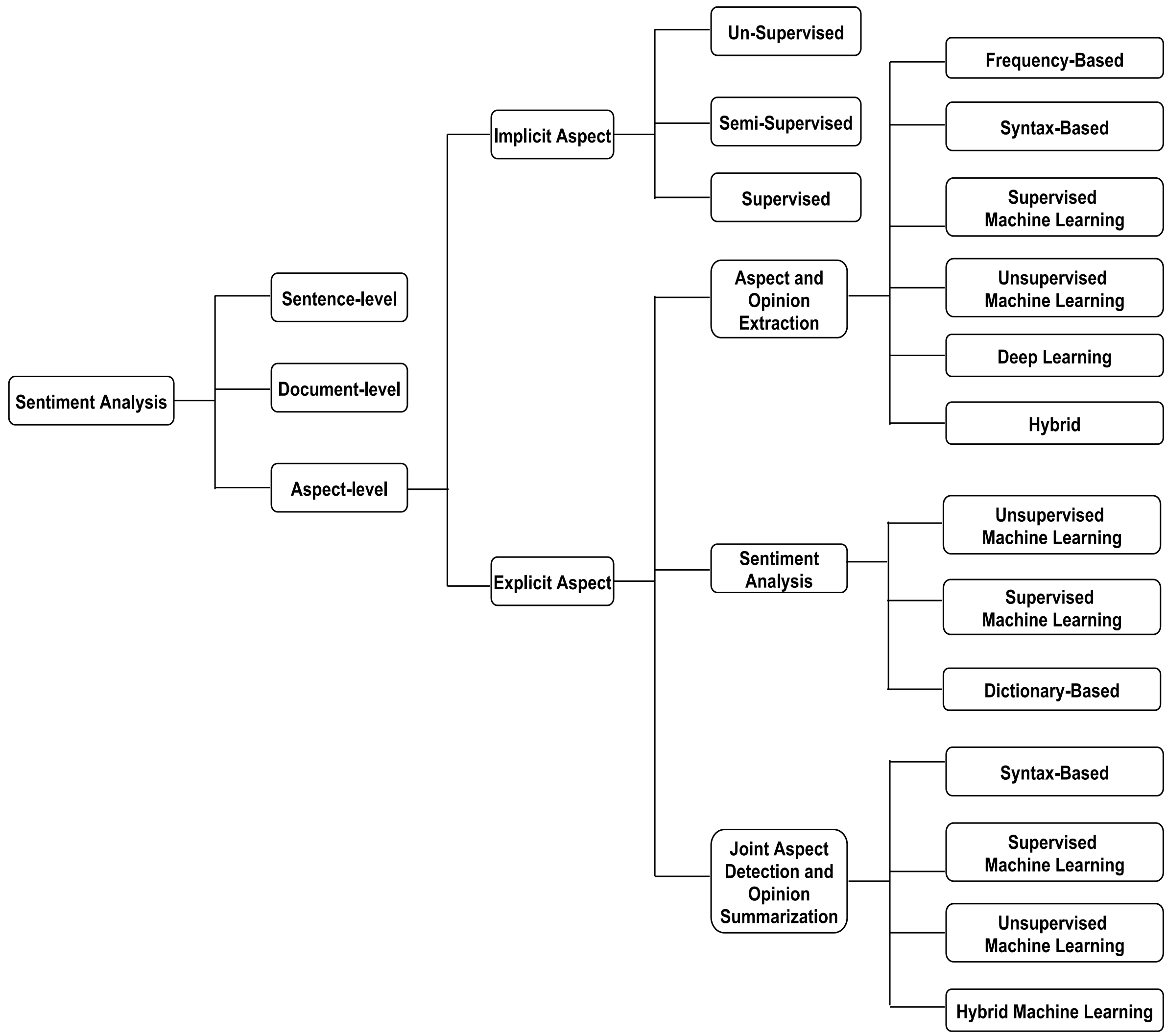

Fig. 3. Aspect-based Sentiment Analysis approaches.

pairs compared to the frequency-based method. A small seed set is a considerable superiority of this method for extracting aspects in place of a huge corpus.

As the power of *supervised* aspect detection methods relies on the features used, feature creation usually comprises additional methods, such as frequency-based methods, for constructing the most prominent features. For example, supervised aspect extraction is a sequence labeling problem that uses a linear Conditional Random Field (CRF). CRF is frequently used in unstructured text mining to test the effectiveness of a sentence's entire word order. CRF automatically considers the word context while assigning it a label. CRF uses numerous features for extracting the appropriate word label, which incorporate the actual phrase, its POS tagging, an immediate relationship between the term and the expressed opinion, the noun phrase holding the word that is adjacent to an opinion expression, and the word being part of a sentence that indeed contains a sentiment expression [11]. Li et al. [26] propose a revolutionary cohesive architecture that integrates implicit sentiment knowledge and explicit syntax knowledge to more efficiently complete all ABSA jobs. In *unsupervised* machine learning, most approaches use the concept of topic model Latent Dirichlet Allocation (LDA). It is assumed that each review sentence is a combination of topics. LDA generates unlabeled topics, which limits a direct connection between aspects and the topics. Although most classification models have their exceptions, some classification models are considered both in supervised and unsupervised machine learning approaches. The hybrid methods integrate supervised and unsupervised approaches, where opinion lexicons are important in most methods. Some of the recent work shows the utility of the hybrid approach. They first extracted the aspects in an unsupervised manner and supplied these aspects as labels for supervised deep learning network training. This survey focuses on deep learning approaches for aspect-level sentiment analysis.

A rigorous feedback analysis at the feature-driven sentiment analysis level requires entities and their related aspects to be extracted [27]. Further, ABSA classifies sentiments correlated with these entities and aspects. Some techniques utilize a predefined set of aspects, whereas the remaining methods randomly obtain aspects from the text [11]. Examples of different aspects are products, restaurants, hotels, services, events, etc. In laptop reviews, the laptop itself is an entity. At the same time, all the characteristics, properties, or attributes related to that product (screen, display, keyboard, etc.) are aspects or features of the laptop. The sentiment (p) and the opinion target (e) can be expressed implicitly or explicitly. Explicit aspects and sentiments are usually mentioned in the text (Fig. 4 represents a taxonomy of explicit aspect extraction techniques). In contrast, it is necessary to extract implicit characteristics from the content utilizing supplementary domain and contextual knowledge. For instance,

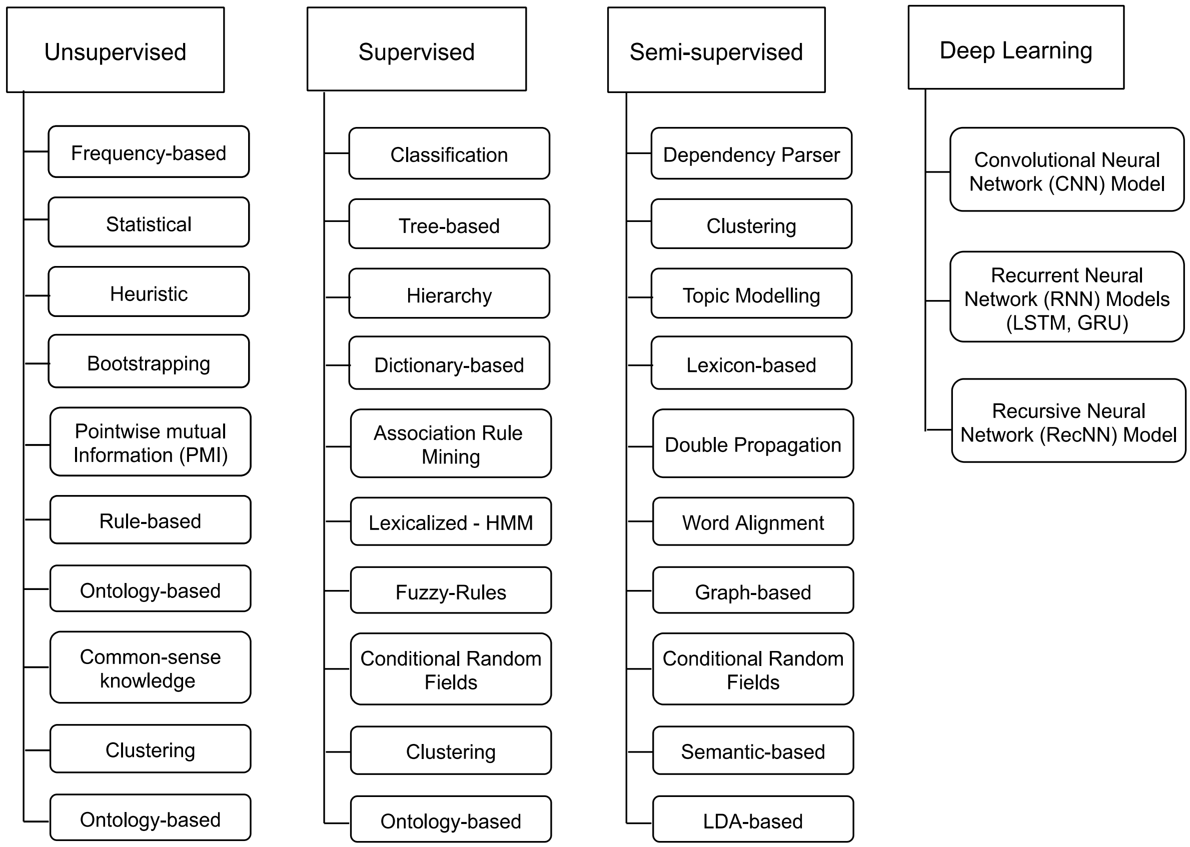

Fig. 4. Taxonomy of explicit aspect extraction techniques.

the phrase, "The pizza of this hotel is fantastic", uses the explicit aspect term "*pizza*" for the entity *Hotel*. On the other hand, "The food in this hotel is expensive" implicitly holds the aspect of *price*. The commonly used aspect extraction and aspect category detection strategies are described in more detail in subsequent parts of this survey.

4.1. An overview of aspect extraction techniques

Rule-based, machine learning, and deep learning methods make up the majority of aspect extraction techniques. In the early phase, Hu and Liu [1] studied aspect extraction from the opinion reviews in aspect-level opinion analysis. In accordance with the dependency links between the nouns, noun phrases, and opinions, they applied a set of rules to determine how frequently they appeared. Only extracting explicit aspect phrases was the main goal of this study. They have applied association rule miners to form aspects that depend on high-occurrence terms in specific categories to work well [28]. Blair-Goldensohn et al. [29], Popescu and Etzioni [30] improved Liu's work to detect single-word and multi-work aspects as an entity's attribute by formulating point-wise mutual information (PMI) for each aspect word. Li et al. [31], calculating *Point-wise Mutual Information* [28,29], language-model-based methods, lexicon-based models, etc., can extract explicit aspects from the review [30]. However, according to [32,33] these methodologies cannot extract implicit characteristics, making it impossible to extract contextual knowledge from the review.

Additionally, the effectiveness of aspect extraction utilizing rule-based techniques improves with the precision of the parsers and taggers. Additionally, these techniques depend on the sentence's linguistic limitations being accurate. The dependency tree has been used and restructured in recent studies to simulate how text's syntactic relationships are formed. Though, the original dependence tree's structural data is destroyed throughout the restructuring process. In order to efficiently locate the pertinent contextual data concerning the aspect word for a sentence's encoded, [34] language-based neural network structure, which performs an in-depth investigation on the syntax graph. Consequently, certain legitimate nouns and noun phrases were not extracted as aspect phrases using rule-based approaches. Next, topic modeling using LDA [17] outperformed linguistic-based models. However, the effectiveness of aspect categorization recognition suffers from an absence of sensible and context-sensitive knowledge [35,36]. Following, the entity's feature extracted using the language model as given in [37] produces a low precision value due to noise in the extracted aspects. Singh Chauhan et al. [18] further improved [25] aspect extraction results on Semeval-16 using the rule-based and similarity-based pruning methods. Table 5 presents the widely used non-deep-learning-based techniques featuring ATE and ACD subtasks of interest to academicians and researchers in the last two decades.

Poria et al. [33] focus on rule-based linguistics, including stop words and negations. The shortcoming of linguistic-based techniques is that they need to be designed manually, and the performance of these rules requires the grammatical accuracy of the review. Methods for supervised learning, such as CRF, SVM [57–59], etc., achieved significant results as sequence labeling subtasks in ABSA [60,61].

Supervised learning methods have shown performance improvement with a sequence labeling problem [57,59]. These supervised methods can establish an association employing contextual details among intra-sentence nouns along with noun words as aspects. However, supervised techniques yield poor results whenever aspect extraction needs the entire review's contexts. Additionally, owing to inter-sentence dependencies in traditional feature development, it is difficult to determine the

Table 5
Summary of non-deep learning aspect extraction approaches.

Article	Domain	Algorithm	Approach	Performance (%)		
				Precision	Recall	F-score
[1]	Product	Frequency-based	Frequency-based	72	80	
[30]	Product	PMI	Unsupervised technique	76.68	77.44	
[38]	Product	Frequency-based	Frequency-based	79.10	82.40	
[39]	Movie	Syntax-based	Syntax-based			52.90
[40]	Restaurant Product	Dictionary-based	Supervised learning	76.60	75.10	
[37]	Product	Frequency-based	Frequency-based	87		
[29]	Restaurant	Hybrid	Supervised machine learning	95	82.20	
[41]	Product (mobile)	NLP + statistical (syntax-based)	Frequency-based			74.07
[42]	Product	Double propagation	Syntax-based	88	83	
[43]	Product	Frequency-based	Hybrid	92.40	62.70	
[44]	Camera	Lexicalized HMM	Supervised learning			82.70
[45]	Movie Product	Skip Tree CRF	Supervised machine learning	86.6 82.6	69.3 76.2	
[46]	Movie	CRF	Supervised learning	74.90	66.10	
[47]	Car Mattress	Double propagation	Syntax-based	78 77	56 64	
[48]	Phone	Bootstrapping	Syntax-based	76	68	
[49]	Cell phone	Association rule mining	Frequency-based	76.29	72.71	74.46
[50]	Product	Dependency parser	Hybrid			76
[51]	Product	Frequency + probability	Unsupervised technique			
[52]	Product	Word alignment + graph-based	Unsupervised technique	78	73	
[53]	Restaurant	Joint modeling	Unsupervised technique	74.30	86.30	
[33]	Product Restaurant Laptop	Rule-based	Unsupervised technique	87.39 85.21 82.15	91.42 88.15 84.32	
[31]	Product	Bootstrapping	Unsupervised technique			
[15]	Product	Topic modeling	Unsupervised technique	67.3	62.4	
[54]	Restaurant	Hierarchy with co-occurrence	Supervised machine learning	94.7	75.8	84.2
[25]	Product	Rule-based + Lexicon	Unsupervised technique	87	92	89
[18]	Restaurant	Rule-based + aspect pruning	Hybrid technique	81.02	77.09	79.01
[55]	Restaurant	SS-LDA	Unsupervised technique	Accuracy: 62.25%		
[56]	IMDB	Fuzzy entropy	Unsupervised technique	Accuracy: 69.3%		

proper review's context, thus impairing the system's capacity to identify the appropriate aspect category [36,61,62], Rana and Cheah [17] have used a manually built list of class labeled patterns to identify aspects. A semi-supervised technique [63] used seed words to extract the topic for aspect term extraction. A common strategy for the aspect extraction task is topic modeling. These methods are founded on two models: (i) LDA [33,64] and (ii) Probabilistic Latent Semantic Analysis (pLSA) [65]. In LDA, the most pertinent topic is found in the document, while pLSA incorporates latent rating information from reviews. Ozyurt and Akcayol [55] propose a novel topic modeling-based approach for short sentence LDA. However, topic models extract the review topics; hence, each noun may not be an aspect term and cannot correlate to the topics [17]. Table 5, based on surveys [8,11,17], further presents a existing conventional supervised and unsupervised techniques for aspect extraction and entity extraction from the review. These surveys also covered the limitations, applications, and new dimensions and directions of research in ABSA [66]. Still, a systematic categorization of aspect extraction and aspect category detection techniques is missing [67].

In recent times, incorporating the semantic significance of the content, deep neural network techniques have improved the aspect retrieval findings from supervised classifiers. Different supervised approaches use deep learning models such as Recurrent Neural Networks (RNN), Convolutional Neural Networks (CNN), Long short-term memory (LSTM), and Recursive Neural Networks (ReNN), which contribute significantly to the ATE problem [36]. The focus of ABSA, in further sections of this paper, is explicit aspect extraction and aspect category detection at the entity or aspect level by applying deep neural approaches. This article also seeks to present recent research contributions in deep learning methods using hybrid, attention-based, and BERT-based methodologies for ABSA rather than explaining conventional findings from the studies done in the past on ABSA.

In next Sections 5–7 presents a flow of the deep learning model to discuss the role of the Deep Neural Network (DNN), various input vectors to the DNN, and sequential training, respectively, for ABSA.

5. ABSA using deep learning

In deep learning (DL)-driven machine learning techniques, artificial neural networks are formulated to understand knowledge representations (features), and different kinds of modules are utilized. In which one layer's output is passed along as input to the next computation layer. DL follows the backpropagation process, in which an objective function's gradient is first computed backward through activation [68]. Thus, it gets converted into numerical value vectors irrespective of the input's type (text, image, or audio). Further, these vectors clustered meaningful translation, sentiment analysis, or classification. Therefore, one needs substantial feature engineering, which requires domain knowledge.

DL methods have observed remarkable improvements and have primarily taken off the requirement for strong contextual information. Deep neural networks (DNNs) imply deep learning models to automatically understand prominent features, which is the main focus of this survey. DNNs are artificial neural networks containing several neuron layers or network processors activated by the weighted computation of neurons or environmental sensors [36,69,70]. As in conventional machine learning

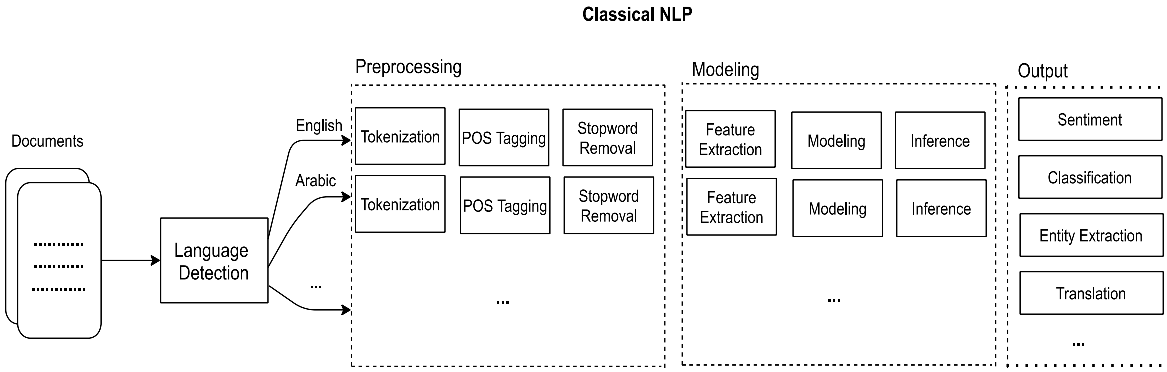

Fig. 5a. Classical NLP using machine learning.

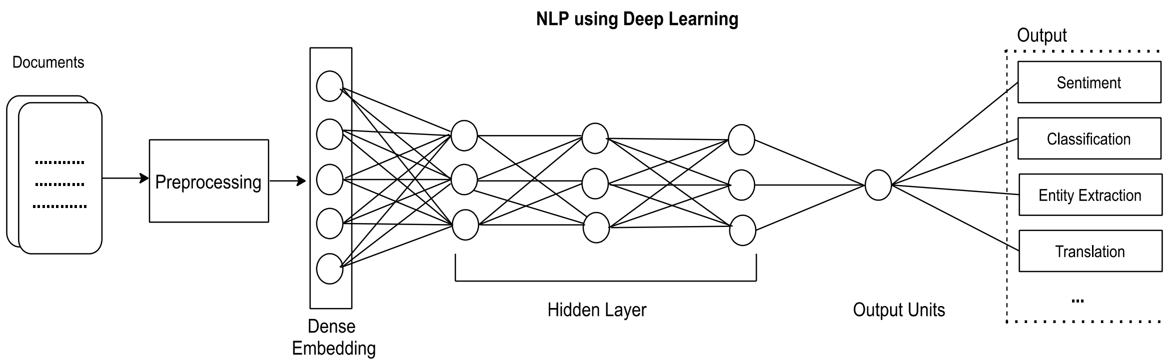

Fig. 5b. Classical NLP using deep learning.

techniques, the sample datasets are split into three categories by the DL approaches: training, testing, and validation data. The training process converts the input into numerical scores in vector form. An error score generated in the training process's initial phase needs to be minimized by training the model with updated weights (parameters) for the target words. These adjustments in the weight parameters are made using the gradient descent method to reduce the error [68].

5.1. Deep neural network for ABSA

Three steps make up DNN approaches for natural language processing (NLP): (1) dense word embedding, (2) multiple hidden layers, and (3) output units, as shown in Fig. 5a and Fig. 5b. The first feature, word embeddings, is a dense numerical vector representation for words that captures their meaning, semantic relationships, and different types of contexts in which they are used. By examining the context in which phrases appear within sentences, the word embedding approach generates word vectors. It subsequently feeds this knowledge into a feed-forward neural network to determine which subsequent words will be clustered adjacently in the vector space [36]. In most DNN-based techniques, word embeddings of the input text are supplied to the hidden layer to overcome manual feature engineering [69,71]. Pre-trained word embeddings are employed in current DNN development to obtain syntactical and semantic information that is deprived of data relevant to the domain [71–73]. Word2Vec [74] is the most frequently employed continuous Bag-of-Words (CBOW) and skip-gram methods of word embedding for encoding context-relevant data. It largely depends on in-domain labeled data because deep learning algorithms employing static word embeddings (word2vec, Glove, etc.) produce an identical embedding for the exact same word across multiple domains.

In the first layer of deep neural networks, only prior knowledge is included in the shallow representations of words. Every time a new task domain is introduced, the network's subsequent layers must be trained from start. The NLP academic community is currently working on methods leveraging transfer learning to close the representational gap adopting an innovative language representation model entitled BERT [75]. Next, multiple hidden layers are designed employing recursive networks, attention mechanisms, feed-forward networks (CNN), recurrent neural networks (LSTM, and GRU), and bi-directional LSTM [36,76]. The non-linear outputs are computed by stacking each hidden layer (formed with multiple neurons). The hidden feature engineering starts with neurons. Let us say $x_1, x_2, \ldots \in R$ is provided with m weight attributes $w_1, w_2, \ldots \in R$, a bias $b \in R$. The activation is calculated as $a = \left(\sum_i^m w_i x_i + b\right)$. Thus, the output function O is:

$$O = act\ (a) = act\left(\sum_i^m w_i x_i + b\right) \tag{4}$$

The hyperbolic tangent ($tanh$), the rectified linear ($RelU$), and the sigmoid (a) are usually always utilized as non-linear activation.

$$\text{sigmoid (a)} = \frac{1}{e^{-a}+1} \tag{5}$$

$$\tanh(a) = \frac{e^{2a}-1}{e^{2a}+1} \tag{6}$$

$$ReLU\ (a) = \max(0, a) \tag{7}$$

The probability distribution p_t of every single input word of the review under the $\boldsymbol{I}$, $\boldsymbol{O}$, or $\boldsymbol{B}$ tagging strategy is generated by a final dense layer. In the range of the tagged distribution p_t and the

Table 6
An example sentence with labels in IOB2 (Aspect-1: Chinese food; Aspect-2: noon).

Words:	Yummy	Chinese	Food	For	Lovely	Noon
Labels:	O	B	I	O	O	B
Type:	Outside	Start aspect	Continuation of aspect	Outside	Outside	Aspect
Aspect:		Aspect-1				Aspect-2

expected tags distribution e_t the categorical cross-entropy loss is minimized as follows:

$$H\left(e_t, p_t\right) = -\sum_{k \in K} e_t\left(k\right) \log\left(p_t\left(k\right)\right) \quad (8)$$

where $K = \{I, O, B\}$, BIO tags set, $e_t(k) \in \{0, 1\}$ and $p_t(k) \in [0, 1]$.

The output units indicate the distribution probability across every label in the final phase. For example, the ith label probability of K class classification with z as a final layer is a softmax function:

$$y_i = softmax(z_i) = \frac{e^{z_i}}{\sum_{k=1}^{K} e^{z_k}} \quad (9)$$

The class label with the maximum probability distribution represents the extracted term. In summary, the DNN uses distributed representation to generalize hidden feature combinations that extend features learned from the training process. Hence, DNN models automatically learn hidden and appropriate features without manual feature engineering compared to conventional machine learning techniques.

In the next sections of this survey, the ATE problem is presented as a task that requires sentence labeling using the inside-outside-begin (IOB2) label as given in Table 6. In the ACD subtask, the label is presented as a binary L = {non-category, category} for predefined categories, attributes, or features. In the sentiment aggregation task, the sentiment is generally a set of four kinds of polarities: P = {positive, negative, conflict, neutral}. In all three subtasks of ABSA, the probability returned from the output units of DNN for assigning a label to each input, where the predicted result is the label with the highest probability. This survey reviews the recent DNN models and their variants.

6. Input vectors to deep neural network for ABSA

The input layer encodes every sentence's individual words into a d-dimensional distribution representation or dense embedded vectors to extract syntactic and semantic information about that word before applying it to the first hidden layer of the DNN [77]. We want these vectors to describe the word's usage, significance, and semantic information in a certain manner.

6.1. Shallow word representation using word embedding vectors

The initial neural probabilistic linguistic-driven word embedding algorithm has been suggested by Bengio et al. [78], where the system searches up the continuous vector leveraging a shared lookup table for the word being searched and its preceding words. Next, the feed-forward network receives a vector to predict the next appropriate word with a probability function.

For example, if a sequence of N words $w_1, w_2, \ldots, w_N$ and past m words are supplied to the aspect extraction model; it predicts the distribution probability of the future words by calculating the parameter Θ which maximizes the objective function O of the model as:

$$\mathrm{O} = \frac{1}{N}\sum_{n=1}^{N} \log\ \mathrm{s}(w_1, w_{n-1}, \ldots, w_{n-m+1}; \Theta) + \mathrm{R}(\Theta) \quad (10)$$

Here, $\mathrm{R}(\Theta)$ represents the regularization and the softmax function estimates $\mathrm{s}\left(w_1, w_{n-1}, \ldots, w_{n-m+1}; \Theta\right)$.

In recent times, word2vec is majorly used as a previously developed word embedding vector by researchers [74]. For training on large input corpus, the word2vec model developed with two variations (skip-gram and CBOW) of the neural network for generating word embedding, as shown in Fig. 6. This dataset (word2vec) is known as Google embedding, made publicly available by the authors. The training data for such 300-dimensional units consists of 100 billion phrases from Google News using the CBOW model. The skip-gram model's loss objective function has been determined as:

$$\mathrm{O} = \frac{1}{N}\sum_{n=1}^{N} \sum_{-m \le j \le m \neq 0} \mathrm{logp}\left(w_{n+j} \middle| w_n\right) \quad (11)$$

where w_n is the centered word and w_{n+j} are neighboring words.

The CBOW is based on the bag-of-words, in which the context of the word obtained from a sequential context. Here, a context window of m words surrounding a word used for the target word w_n at time step n.

$$O = \frac{1}{N}\sum_{n=1}^{N} \log \mathrm{p}(w_n, w_{n-m}, \ldots, w_{n-1}, w_{n+1,\ldots,} w_{n+m}) \quad (12)$$

In ABSA, [71,73] the CBOW framework of Mikolov et al. [74] was trained using an extensive dataset of Amazon product reviews created by McAuley and Leskovec [79]. The Amazon embed vector has 2,441,053 reviews of Amazon items totaling 34,686,770 words (4.7 billion words) spanning June 1995 through March 2013. This model contains 50-dimensional and 300-dimensional word embeddings. In addition, this model includes sentiment-effective text (trained using Amazon product reviews), which was absent in Google embedded vector developed on the Google News corpora. Apart from word2vec models, other models have been developed for word embedding training like Glove (based on Wikipedia and Twitter) [80] or fastText, developed by Stanford University and Facebook, respectively. The fastText learned the word vectors for 157 languages [81]. Besides, 50-dimensional word vectors based on the Wikipedia corpus SENNA were developed [82]. Recently, some researchers have used some variations of 50-dimensional sentiment-specific word embedding (SSWE), such as $\mathrm{SSWE_{h}}$, $\mathrm{SSWE_{r}}$, $\mathrm{SSWE_{u}}$, etc., based on the Twitter text [83].

The initialization of word embedding can be done in several ways, such as random initialization or pre-trained embedding [76]. The random initialization method may encounter in the local minima, stochastic gradient descent. In contrast, the deep learning model's feature learning capability cannot automatically be exploited if fine-tuning the pre-trained model is not done [71]. The experimental results of recent deep learning-based aspect extraction and aspect category detection techniques such as Wu et al. [61],Poria et al. [73],Che et al. [84],Jebbara and Cimiano [85], etc. have shown the effectiveness of fine-tuned pre-trained word embedding. In recent deep learning approaches for ABSA, most researchers have used pre-trained embedded vectors and then fine-tuned them to initialize the word embedding correctly. Moreover, the effective execution of DNN models has shown the importance of pre-trained word embeddings for a large corpus of similar domains [86]. According to the results of the ABSA task [61,73,87], a word embedding scheme with more

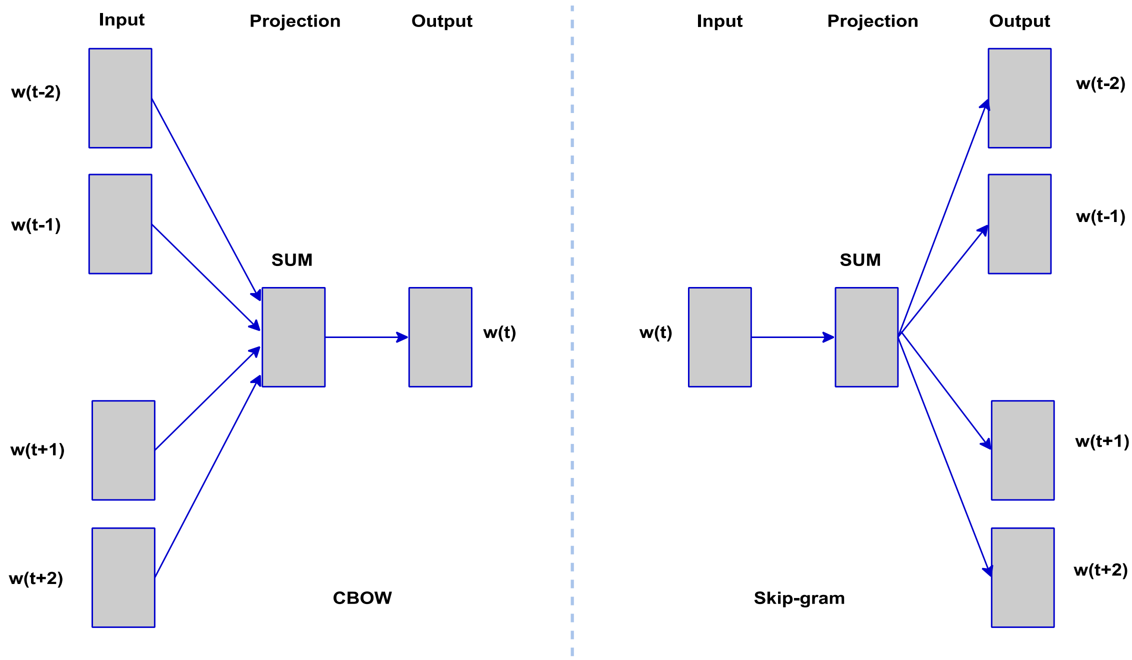

Fig. 6. CBOW and Skip-gram model [74].

sentiment-specific words (Amazon embedding on product reviews) performs better than the word embedding of formal texts (Google Embedding on Google News and Wikipedia). Wang et al. [88] proposed extended dependency-based embeddings created in RNNs which combine the benefits of conventional characteristics and word vectors [61]. A concatenated word embedding with POS is supplied into the hybrid attention model as input [18].

6.2. BERT as embedding layer for aspect extraction

The word2vec, Glove, fastText, and other deep learning algorithms that use static embeddings produce an identical embedding for the same term across a wide range of contexts; as a result, they largely depend on in-domain labeled data. More recently, two pre-trained language models, such as the feature-based method called ELMo [89] and the fine-tuned approach called Generative pre-trained transformer (OpenAI GPT) [90], have demonstrated performance improvement in NLP tasks. The above techniques reduce the work required to set up deep learning models to gather contextual data for obtaining domain expertise. However, these two techniques, particularly with regard to fine-tuned methodology, prevent fully utilizing the power of pre-trained representation. The main problem is that current modeling techniques for languages are unidirectional in nature which limits the system's framework that can be utilized for pre-training the model. A new language representation model called BERT has recently been created by the NLP scientific community to address the discrepancy in the way data is represented. BERT is made to collectively influence the left and right circumstances in every layer in order to prepare deep bidirectional representations derived from unmarked review. Because of this, the pre-trained BERT model can be improved using merely one extra output layer to produce cutting-edge models for a variety of scenarios [75].

Recently, in BERT-based contextual embedding, first, pre-training the general-purpose language representation neural network model with the help of the vast amount of unannotated data on a known domain. The model that was trained is then used as the foundation for an entirely novel job-specific model, which is then fine-tuned. Although the phrases in the review have different discussing contexts, there may be ambiguity in the explanations for multi-word subjects [86]. For instance, the phrases "The story telling is unpredictable" and "unpredictable driving" both include the word "unpredictable" in their sentences. Although the expression unpredictable has a favorable connotation in the literature world, it is detrimental to the car element. Within the initial layer of DNNs, only prior knowledge is included in the shallower representations of phrases. The following layers of the system are consistently taught independently for the various domains [75]. BERT incorporates the POS features of the context into the concept representation after training its context in a domain-specific corpus.

The syntactical and semantic aspects of the textual review are encoded by the BERT using a multi-layer bidirectional transformer encoder rather than a Bi-LSTM. This illustration concurrently addresses the left and right sides of the word's contextual meaning. A completely linked layer and a self-attention layer are both present in the transformer block. Reviews are input into the embedding layer of BERT, which uses the contextual information to produce the token-level representation token, location, and segment embedding are added to determine the input embedding for a certain input sentence. The position embedding indicates the token's position in the review, whereas the token embedding is a unique embedding for each token. All of the sentence tokens' segment embeddings, however, will be the identical and point to the specific sentence to which the token belongs. The acquired key characteristics were utilized in this paper for the aspect extraction method.

Consider [19] the following example is a token sequence for a source of data sentence $x = \{t_1, t_2, t_3, \ldots, t_T\}$, the formulation of the input embedding is as $x' = \{x'_1, x'_2, x'_3, \ldots, x'_T\}$, where x'_i is the combination of token embedding, position embedding, and segment embedding for each input token t'_i; and T is the number of tokens in the token sequence. Next, x' passes with L(12 in our experiments) multi-layer transformer blocks to calculate the corresponding contextual representation as $h' = \{h'_1, h'_2, h'_3, \ldots, h'_T\}$. Finally, this contextual representation of the input token sequence (hidden states (768) of the last layer of BERT) is used as the input features of the sentence for aspect extraction using deep neural networks [75,91]. The input representation of the

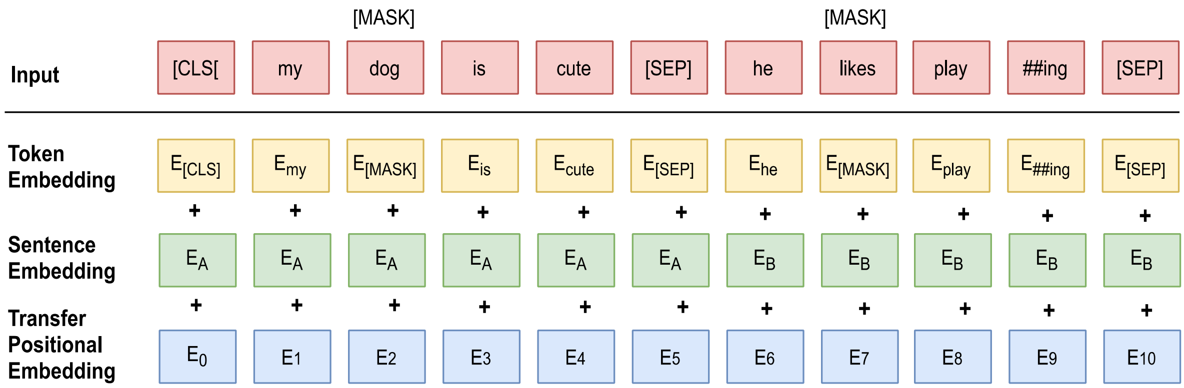

Fig. 7. BERT model taken from [75].

words using the BERT-based technique is shown in Fig. 7 from [75]. In ABSA, dynamic word embedding with (combined token embedding, position embedding, and segment embedding) using BERT [91] used as an input feature in aspect extraction and aspect category detection subtasks. Regarding the distributed illustration of review in the ABSA task, the contextual word embedding leverages the BERT model as a foundation. BERT is helpful to increase the performance of the ATE task from multiple sentences of the review.

The BERT uses subsidiary sentiment knowledge by inculcating sentiment context information into the language representation model [92]. The contextual word embedding of this linguistic appearance model encompassed the inter-sentence relationships between the phrases in a text to comprehend the whole review's context by retaining the semantic significance of the words in a range of domains.

In recent years, researchers [93–95] have applied BERT language representation model to embed the input reviews in order to gather more comprehensible knowledge. Su et al. [96],Wang et al. [97] have greatly enhanced domain-dependent aspect extraction achievements using contextual information. Recently, Zhao and Yu [92] have proven the utility of BERT incorporating domain knowledge into contextual embedding vectors as the language representation model to achieve remarkable performance even with small-sized training data. The impact of BERT as contextual embedding in the language representation model for deep learning aspect extraction methods is discussed in Section 8.3.

6.3. Features for ABSA

Most deep learning aspect extraction approaches have used word embedding (given in Sections 7.1–7.3) as features for the network. The existing DNN based aspect extraction model depends on the language model without requiring manual feature engineering or dependency on a parser or positional features [98]. Moreover, the performance of hidden and multi-word aspect extraction, which requires a contextual understanding of the review, uses the embedding of Word is combined with a number of other feature vectors and supplied into the DNN model's hidden layer. The POS, NP chunks, common-sense knowledge, ontology, etc., are the most common features concatenated with word embedding described below.

POS Tagging Generally, aspect terms are nouns, noun phrases, or NP chunks; thus, POS is an essential feature in ATE [1]. The POS tagging has varieties of classifications, such as six tags given with Stanford Tagger in [73], four tags with Stanford Tagger in [71], 45 tags according to Penn Tagger in [85]. The four types of tags based on Stanford Tagger are adjectives, adverbs, nouns, and verbs. According to [71], the five categories of chunks are noun phrases (NP chunk), adverb phrases (ADVP chunk), verb phrases (VP chunk), adjective phrases (ADJP chunk), and prepositional phrases (PP chunk). The n tags (n part of speech) encodes n-dimensional binary vectors. The two feature vectors (POS and word embedding) are further concatenated and supplied to the DNN model for ATE. The findings of the assessment results also justify the effectiveness of POS and word embedding as combined features for hidden layers of the deep learning models.

In both the subtasks ATE, and ACD, common-sense knowledge such as SenticNet has proven to be another essential feature [85,99]. They incorporate conceptual and contextual information to associate the context and multi-word aspect terms from the multi-sentence review. Besides, the accuracy of the latest ABSA techniques depends on the performance of the dependency parser and POS tagger. The SenticNet includes over 50,000 concepts with essential properties for domain-specific aspect extraction. For example, according to [99], the concept "pizza" has the associated property "kindOf-food" which has a relation with the "restaurant" domain and the aspect category "*FOOD#QUALITY*". Therefore, aspect extraction with Sentic LSTM generates better accuracy than the baseline aspect extraction models [99].

7. Training the sequential data on DNN model for aspect extraction

Generally, the backpropagation process used to train a neural network involves backward gradient computation for each parameter and then updating with stochastic gradient descent [36,100]. If $x_1, x_2, \ldots, x_n$ is the input, $y_1, y_2, \ldots, y_n$ is the deep learning model's output and the authenticate labels are $\bar{y}_1, \bar{y}_2, \ldots, \bar{y}_n$, the motivation behind this is to calculate the function $y = f(x)$ that estimates the loss $\mathcal{L}$ as the cost function in relation to the Θ is:

$$\mathcal{L}(\Theta) = \frac{1}{n}\sum_{i=1}^{n} \mathcal{L}(\mathrm{f}(\mathrm{x_i};\Theta), \bar{\mathrm{y}}_i) \tag{13}$$

Next, in the loss-reducing procedure, technique also includes function $\mathrm{R}(\Theta)$ to overcome overfitting issues using Θ' as:

$$\mathcal{L} = \mathrm{argmin}_{\Theta} \frac{1}{n}\sum_{i=1}^{n} \mathcal{L}(\mathrm{f}(\mathrm{x_i};\Theta), \bar{\mathrm{y}}_i) + \lambda \mathrm{R}(\Theta) \tag{14}$$

In ABSA, the categorical cross-entropy loss is used for label prediction and distribution for tagging aspects [36].

$$\mathcal{L}(y, \bar{y}) = -\sum_{i} \bar{\mathrm{y}}\ \log(y_i) \tag{15}$$

From time to time, SVM and CRF may be used as non-neural classifiers for accomplishing higher performance ([72,101,102],

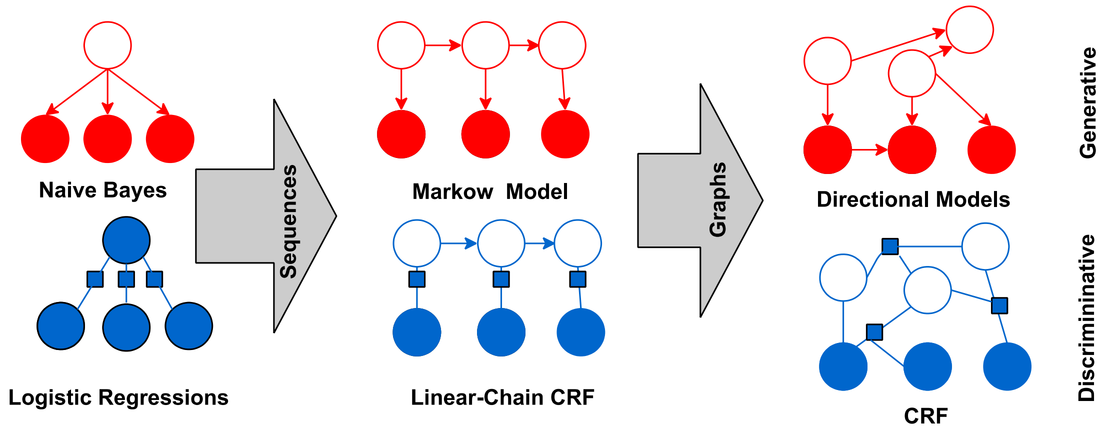

Fig. 8. CRF with other models given by Sutton et al. [103].

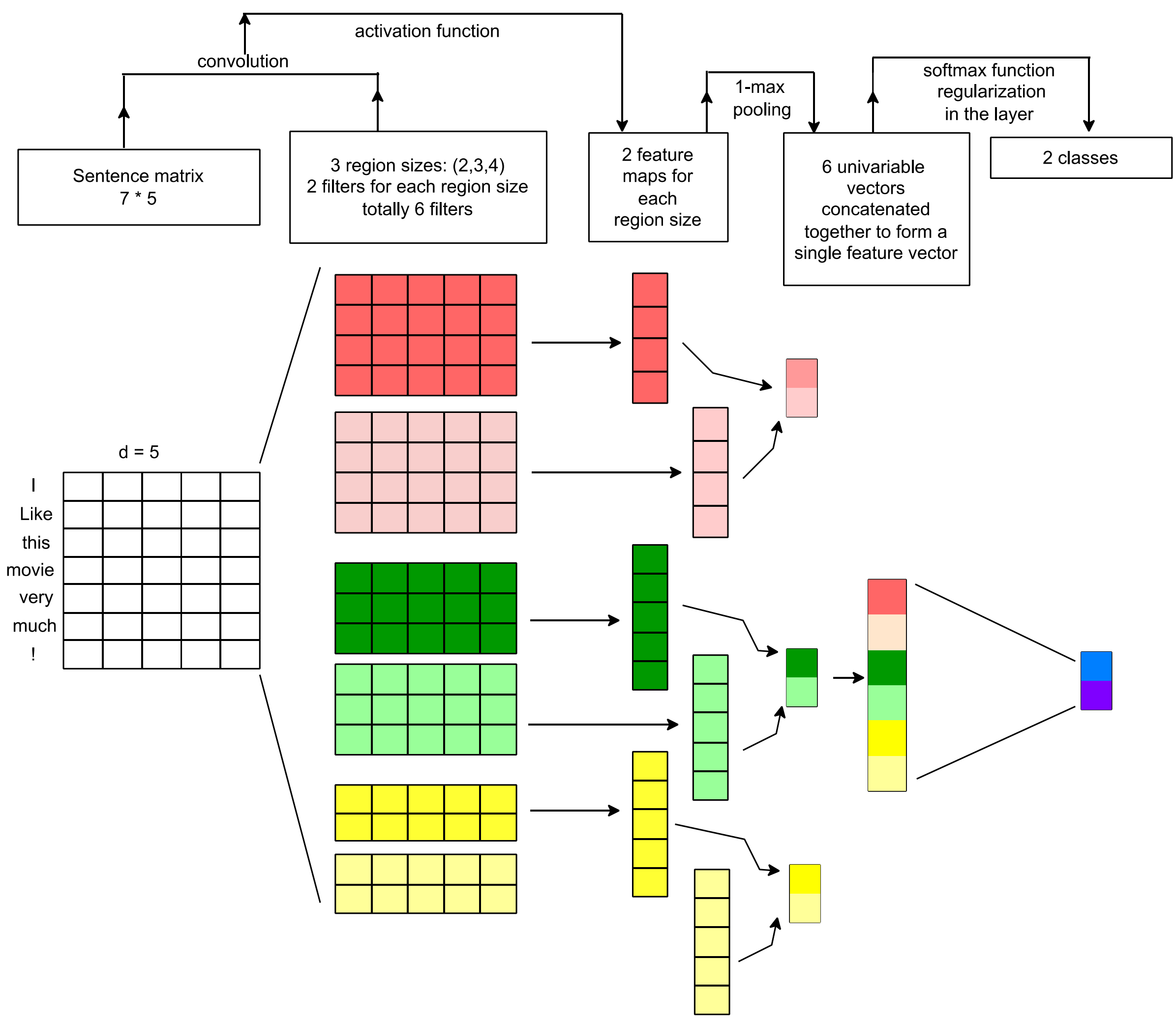

Fig. 9. CNN with 4 layers for NLP tasks.

Zhang et al., 2018). The whole phrase pattern is used by CRF to determine the likelihood of an aspect being labeled. Fig. 8 shows the CRF with other models given by Sutton et al. [103].

7.1. Convolutional neural network CNN) for aspect extraction

Collobert et al. [82], Kim [104] recommended the importance of CNN models in a variety of NLP-based applications. In the conventional single-layer CNN model for NLP applications, a sequence labeling task contains four layers: an input layer, a convolutional layer, a max-pooling layer, and a extensively connected layer, as given in Fig. 9. Each CNN layer model is represented below:

First, the *Input layer* represents the input review of length m as follows:

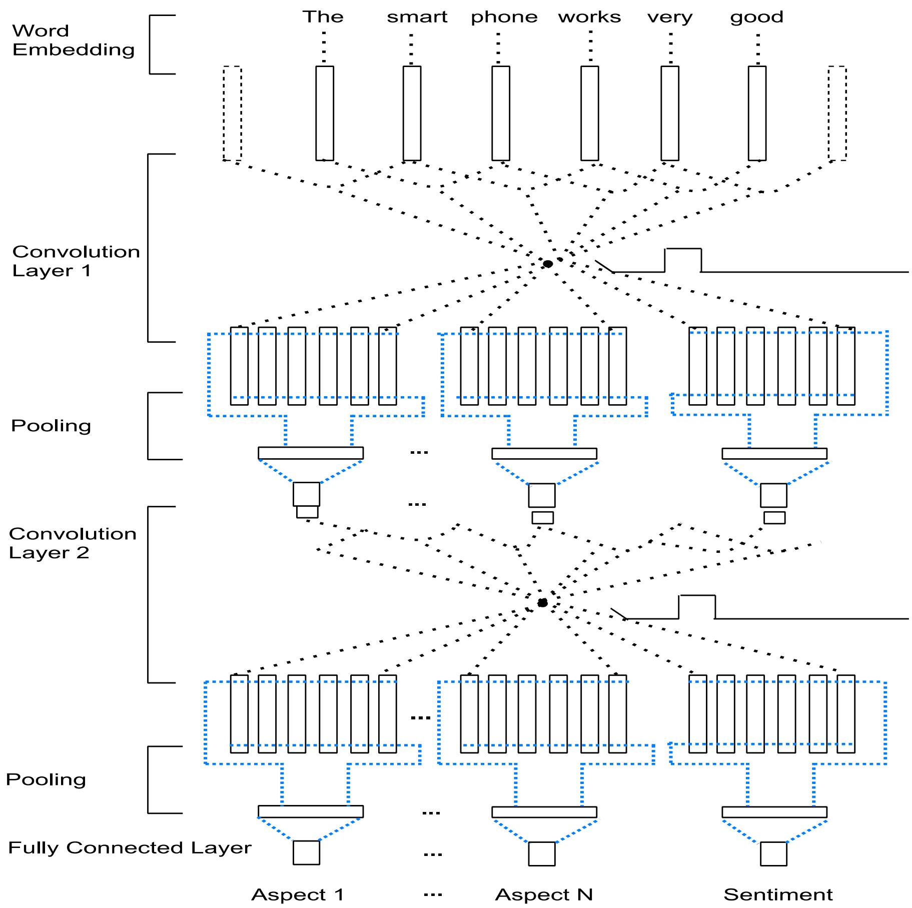

Fig. 10. Aspect-based sentiment analysis using CNN model from [105].

$x_{1\ to\ m} = x_1 \oplus x_2 \oplus \ldots x_m$; here x_i belongs to the d-dimension word embedded vectors (explained in Section 6) concerning the ith word of the input sentence, whereas $\oplus$ represents the concatenation operation. The $m \times k$ representation of reviews having dynamic or static streams of input are provided into the following Convolutional layer.

The second layer; the *Convolutional layer*, contains multiple filter widths and feature maps, extracts the new features C_i with the filter $f \epsilon \mathbb{R}^{nk}$; where n-words denotes the window from i to $i + n - 1$ using the non-linear activation functions like *sigmoid*, *tanh*, or *ReLU* (explained in Section 5.1) as

$C = a(f.x_{i\ to\ i+n-1} + bias)$; where bias $\epsilon \mathbb{R}$ and a is activation function.

Thus, the feature map can be represented as

$$C = [C_1, C_2, \ldots, C_{m-n+1}] \tag{16}$$

Next, the *max-pooling layer* picks the feature map with the maximum value from C calculated above in the second layer. The selected feature map ($C' = \max(C)$) is the extracted feature concerning each filter. Finally, the feature vector with r filters f formulated as $F = [C'_1, C'_2, \ldots, C'_r]$. The final softmax layer output to identify valid aspects obtained using the softmax function is:

$$y_i = softmax\,(WF + bias) = \frac{e^{WF+bias_i}}{\sum_{k=1}^{K} e^{WF+bias_k}} \tag{17}$$

Here, the output y_i of the softmax layer with k classes is the overall label's probability distribution. Earlier this decade, CNN-based DL models significantly contributed to ABSA [73,106]. The primary capability of CNN is to bring out the most relevant n-gram features from the textual sentence to form a semantic representation. Next, this semantic representation is used for several distinct categories of tasks, including the aspect extraction subtask ([69,98], Zhang et al. 2018). In the aspect extraction task, the anticipated aspect phrase is acted upon by the label having the optimum probability. In ATE using the CNN model, the Convolutional layer may retrieve the most prominent features (n-gram) as aspect terms corresponding to a particular feature map. On the other hand, the max-pooling layer creates an output with a fixed length regardless of the filter window's length n. Fig. 10 represents the ABSA using the CNN model from [105].

7.2. Recurrent neural network (RNN) model for aspect extraction

The basic RNN model follows the Elman network [107], consisting of hidden connections with direct cycles. The model's hidden states are all dependent on earlier hidden states, current input, similar parameters, and the activation function for every recurrent unit [108], as shown in Fig. 11. In this network, if x_t is the input at a time step t, then the hidden state s_t will be calculated as:

$$s_t = a\,(Ws_{t-1}Ux_t) \tag{18}$$

where W represents the weighted matrix activation connecting x_t and s_{t-1}, a is non-linear activation function (explained in Section 5.1).

The RNN model's output is calculated as follows:

$$o_t = softmax(Vs_t) \tag{19}$$

Compared to conventional feedforward networks, the additional characteristics of the RNN model are that (1) learning of model needs a smaller number of parameters. All the parameters in

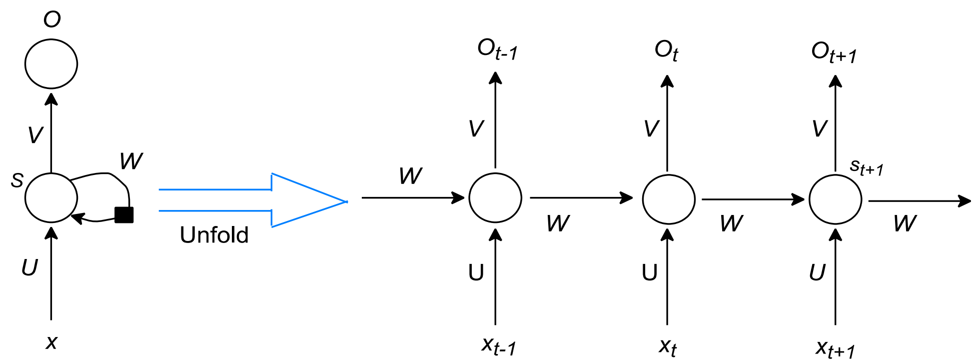

Fig. 11. The basic Recurrent Neural Network (RNN) Model [108].

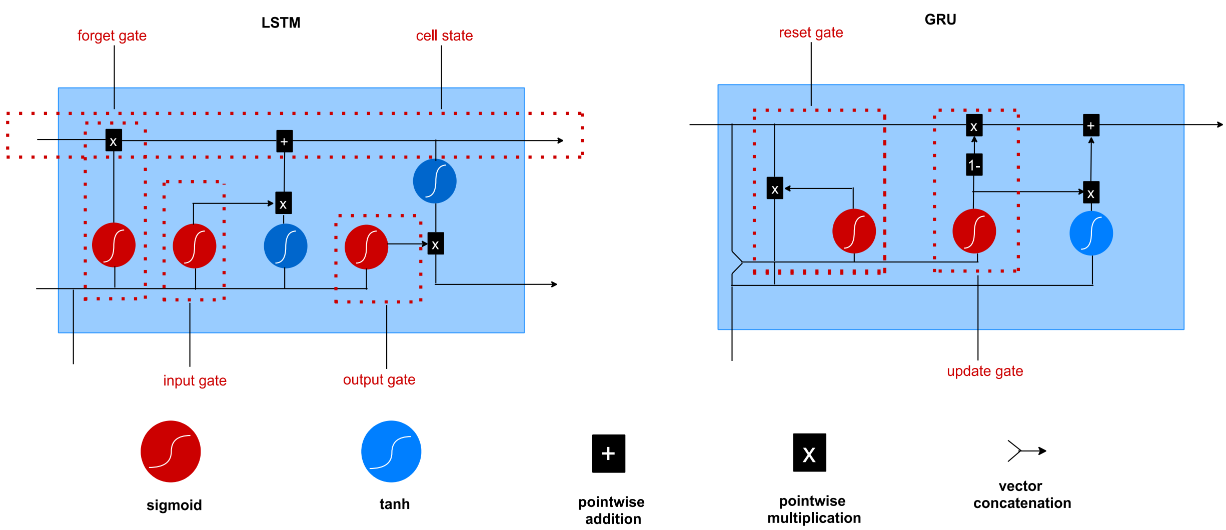

Fig. 12. Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) Models [109].

each recurrent unit are the same (CNN uses a different parameter for each step); (2) effectively processes the sequential data using the output of previous computations (memory) (Zhang et al., 2018). However, the backpropagation process of RNN may encounter issues with disappearing gradients (extremely low gradients) or difficulties with exploding gradients (the gradient rises too much). These problems, in turn, cause an unstable learning process and interrupt the parameter fine-tuning to increase the gradient. These limitations further restrict applying the RNN model (Fig. 11) to long sequences (sentences or reviews) [76]. These weaknesses of RNN have been overcome with variations of RNN like LSTM and GRU [109], according to Fig. 12.

The memory cell (cell state) is the core basis of LSTM, which uses three gates: forget gates (f), input gates (i) and, output gates (o) to control the read, write, and update operations of the internal cell state. Suppose the input is x_t at time step t, and previous state output is s_{t-1}, the forget gate decide, when upgrading the memory cell, which details should be ignored and what data should be retained. W, U is the weighted matrix function, $\odot$ denotes element-wise multiplication. The computation of hidden state (s_t) for LSTM network is given as:

Input gate:

$$i_t = sigmiod\left(W_i s_{t-1}, x_t + bias_i\right) \tag{20}$$

Forget gate:

$$f_t = sigmiod\left(W_f s_{t-1}, x_t + bias_f\right) \tag{21}$$

Output gate:

$$o_t = sigmiod\left(W_o s_{t-1}, x_t + bias_o\right) \tag{22}$$

New Memory Cell:

$$C_t' = tanh\left(W_c s_{t-1}, x_t + bias_c\right) \tag{23}$$

Final Memory Cell:

$$c_t = i_t \odot C_t' + f_t \odot c_{t-1} \tag{24}$$

Hidden state:

$$s_t = o_t \odot \tanh\left(c_t\right) \tag{25}$$

The GRU only contains reset gates (r) and the update gate (u); which regulate the flow of information as the LSTM does without the memory cell. The computation of hidden state (s_t) for GRU network is given as:

Update gate:

$$u_t = sigmiod\left(W_u\left[x_t\right] + U_u s_{t-1} + bias_u\right) \tag{26}$$

Reset gate:

$$r_t = sigmiod\left(W_r\left[x_t\right] + U_r s_{t-1} + bias_r\right) \tag{27}$$

New Memory content:

$$s_t' = tanh(W[x_t]) + U\left(r_t \odot s_{t-1}, x_t\right) + bias \tag{28}$$

Hidden state:

$$s_t = u_t \odot s_{t-1} + (1 - u_t) \odot s_t' \tag{29}$$

Based on the basic RNN model and its variant, LSTM, GRU predicts the next word in the sequence using past words. ABSA, along

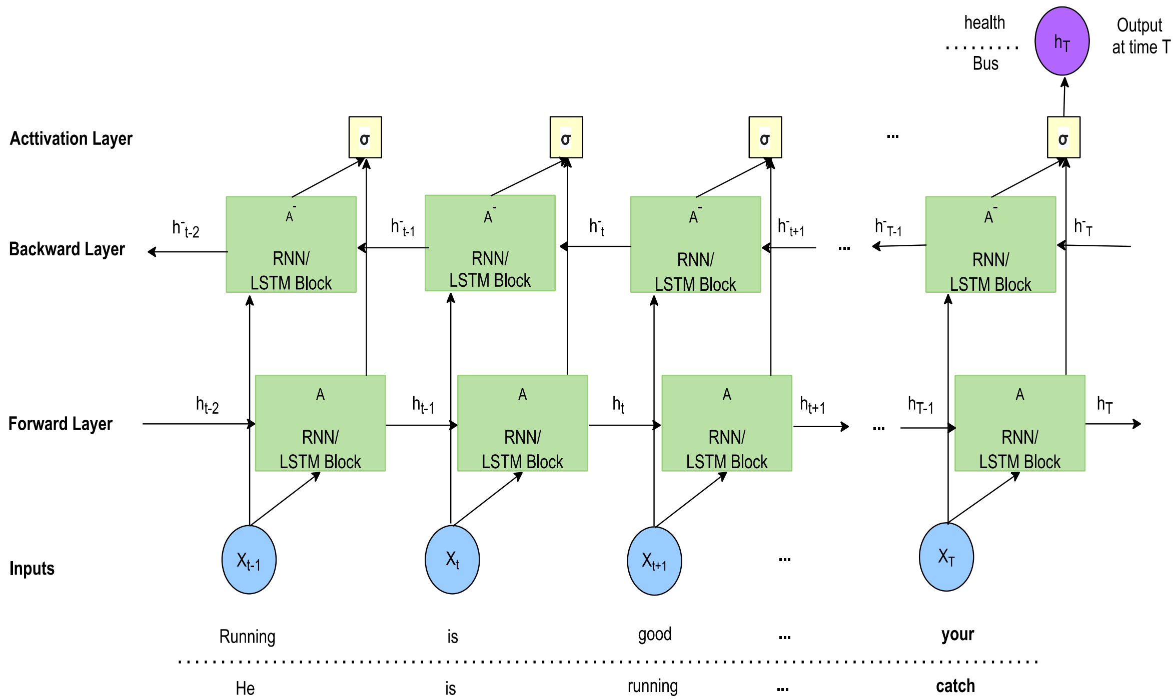

Fig. 13. Bi-RNN Model containing forward and backward hidden layers.

with many other applications, requires complete contextual information about the sentence or review to make the prediction. Therefore, both past and future words need to be read in parallel to understand the sentence's meaning. To extract a word's previous and subsequent tokens in the phrase provide information, Bidirectional-RNN (Bi-RNN) incorporates a backward layer as well as a forward layer [81,110]. Fig. 13 shows that in the Bi-RNN (Bi-LSTM) model, at the time step t, the current input x_t and past hidden state ($h^{\leftarrow}_{t-1}$) compute the forward hidden layer ($h^{\rightarrow}$); whereas the current input x_t and next hidden state ($h^{\rightarrow}_{t-1}$) used to calculate the backward hidden layer ($h^{\leftarrow}$). Concatenating the contextual representation into the lengthy sequence vector as seen below yields the final hidden state:

$$h_t^{\leftrightarrow} = \left[h_t^{\rightarrow} : h_t^{\leftarrow}\right] \tag{30}$$

Recurrent Neural Networks have been the preferred choice of researchers and academicians in different subtasks of ABSA in the past few years. The fundamental concept of RNN is that a sentence or document (sequence) can be represented using a fixed-size vector and then fed all the tokens into the recurrent unit to capture the sequential nature of the textual review [36,76,111]. RNN-based aspect extraction models adopt customizable executing stages, unlike CNN models, where subsequent RNN's output is reliant on prior computations. In order to comprehend the context of long phrases and multi-sentence reviews, RNN models might thus be used [77].

The n-gram feature selection out of the input reviews is the primary addition to the CNN models. However, RNN-based techniques are primarily preferred in ABSA. In RNN models, hidden sequential patterns capture recurrent units, making the computation step more flexible than CNN. Moreover, ABSA using RNN techniques can capture contextual dependencies in sentences. However, the vanishing gradient and exploding gradient problems of RNN-based aspect extraction improve with the use of LSTM and GRU [109].

7.3. Attention mechanism for aspect extraction

In the present day, each word of the review does not uniformly participate in the context extraction subtask. Conventional RNN-based models can examine unrelated information significantly when relating the context to the appropriate subject matter for the sentiment that is expressed in the review. On the other hand, LSTM and deep learning models comply with linear organization with the capacity to give greater importance to terms that are crucial yet extract wrong information. One way to capture the noteworthy influence of the phrases on the tone of the review is to use the attention mechanism, which forms a dense vector while considering the weights of the different word vectors [32]. Within this system, the model preconceives a particular chunk of the review rather than encrypting knowledge into a vector of certain lengths. Fig. 14 presents the decoder-encoder-based [113] traditional attention-based Bi-RNN model [112]. Deep learning with an attention mechanism can analyze prominent linguistics and seize the remarkable impact of the words on the sentiment of the review sentence. Additionally, the domain knowledge is properly extracted to set up a long-term dependency among precise words.

Suppose, the given reviews are $S = \{X_1, X_2, X_3, \ldots, X_T\}$ at a time t, the output y_t (BIO tag set) represents encoder states $H = \{h_1, h_2, \ldots, h_T\}$ and decoder state s_t [24].

$$\boldsymbol{h}_t = biLSTM(h_{t-1}x_t; \Theta_{encode}), \tag{31}$$

Here $\boldsymbol{h}_t$ can be represented as given in (Eq. (31)), where $\boldsymbol{h}_t^{\rightarrow}$ and $\boldsymbol{h}_t^{\leftarrow}$ represents forward and backward LSTM network output, respectively.

$$\boldsymbol{h}_t = \boldsymbol{h}_t^{\rightarrow} \oplus \boldsymbol{h}_t^{\leftarrow} = biLSTM^{\rightarrow}(\boldsymbol{h}_{t-1}x_t; \Theta_{encode}) \oplus biLSTM^{\leftarrow}(\boldsymbol{h}_{t-1}x_t; \Theta_{encode}) \tag{32}$$

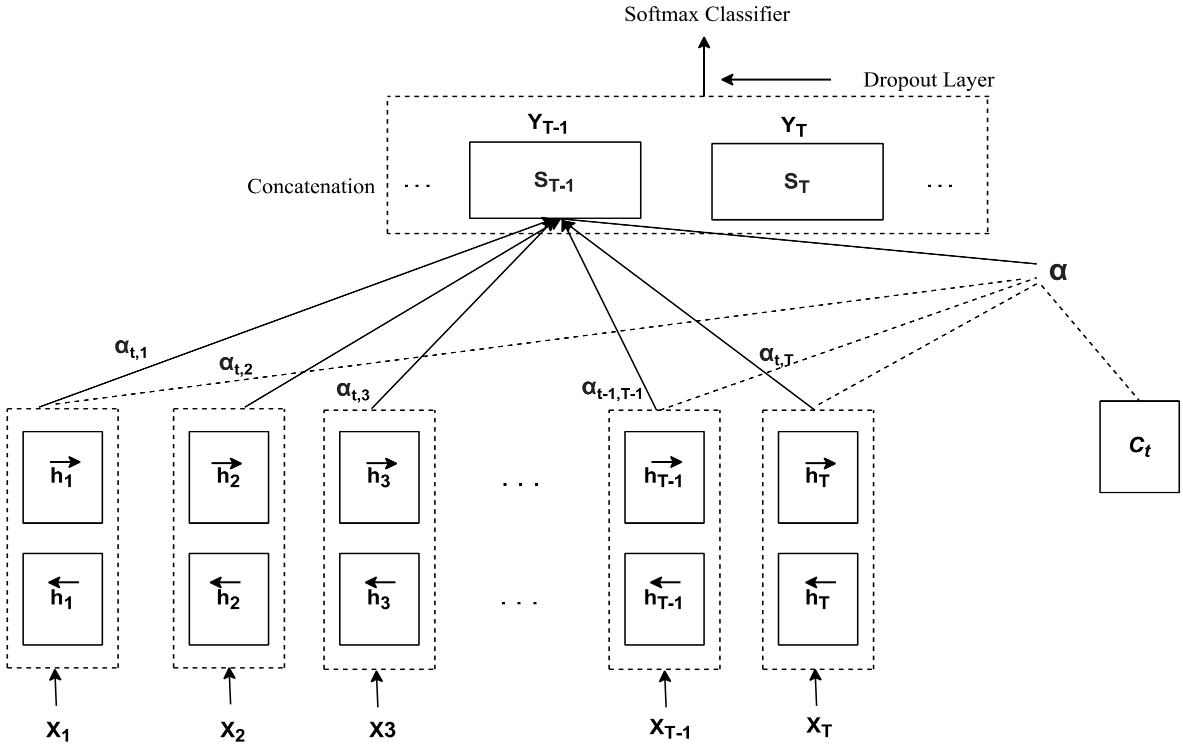

Fig. 14. Attention-based Bi-RNN model taken from [112].

The computation of s_t given in Eq. (32) is as:

$$s_t = f(s_{t-1}, y_{t-1}, c_t) \tag{33}$$

The decoder produces the Y according to X to forecast the following label y_t using context vector (c_t) and previous outputs $[y_1, y_2, y_3, \ldots, y_{t-1}]$. Attention weight $\alpha_{ti} = \{\alpha_{t1}, \alpha_{t2}, \ldots, \alpha_{tT}\}$ will be the factor used for determining the context vector (c_t) as:

$$c_t = \sum_{i=1}^{T} a_{ti} h_i \tag{34}$$

A feed-forward network bi-LSTM is used to calculate the e_{ti} employing inputs such as s_{t-1} and h_i:

$$e_{ti} = biLSTM(s_{t-1}, h_i) \tag{35}$$

In detail the attention e_{ti} computed as:

$$e_{ti} = ev_{aspect} tanh(Us_{t-1}, Wh_i) \tag{36}$$

Hence, following that, the likelihood of the resultant labeling defined as:

$$P\left(Y|X\right) = \pi_{t=1}^{T} P\left(y_t | y\left[1{:}t-1\right], c_t\right) \tag{37}$$

where $y\left[1{:}t-1\right]$ is equal to $[y_1, y_2, y_3, \ldots, y_{t-1}]$, and the probability $P(y_t|y\left[1{:}t-1\right], c_t)$ defined as α_{ti}.

The probabilistic distributions q_t of every input term of the review using ***I***, ***O***, or ***B*** tagging strategy is generated by the last dense layer. The following shows the variation of the labeled distribution q_t and the anticipated tag distribution p_t for optimization of categorical cross-entropy loss.

$$H\left(p_t, q_t\right) = -\sum_{k\,\in\,K} p_t\left(k\right) \log\left(q_t\left(k\right)\right) \tag{38}$$

where $K = \{I, O, B\}$, a set of IOB2 tags, $p_t\left(k\right) \in \{0, 1\}$ and $q_t\left(k\right) \in [0, 1]$.

In the strategies suggested, the attention weight a_{ti} is calculated using a softmax function using the ith label likelihood of the K class classification as:

$$\alpha_{ti} = P\left(y_t = i|X, \Theta\right) = softmax\left(e_{ti}\right) = \frac{\exp\left(\mathrm{e}_{\mathrm{ti}}\right)}{\sum_{\mathrm{i}=1}^{\mathrm{T}} \exp\left(\mathrm{e}_{\mathrm{ti}}\right)} \tag{39}$$

here e_{ti} = a(s_{t-1}, h_i) is used for attention energies

In addition, it is assumed that, given the aspects and their aspect categories and aspect terms, the context vector generated in the attention mechanism is most favorable to extracting aspects by targeting specific sections of the sentence [24]. Further, the currently used ABSA methods solely record the semantic significance of the phrases., thus being incapable to recognize dependencies for noun phrases across the review sentences, which extract some invalid aspect terms. However, latest ATE methods using deep learning work significantly after being trained on a larger number of samples. Still, in the last few years, the ability of the attention system to acquire contextual information from the reviews for better aspect prediction (Zhang et al., 2018).

In recent times, multiple variants of attention mechanism-based aspect extraction have significantly improved the performance of ABSA on the SemEval datasets for the restaurant and laptop reviews. Several well-known attention-based strategies are coupled multilayer attention based on GRU [114], a residual attention network [115], Co-attention LSTM [116], BERT-based co-attention [117], a multi-domain attention model [118], self-attention with the BERT model [96], and an augmented knowledge sequence-attention mechanism [119]. In Section 8, a discussion on CNN, RNN, and its variants for aspect extraction is presented.

8. Discussion on CNN, RNN and its variants for ATE and ACD subtasks of ABSA

This section discusses different CNN and RNN models presented with the article, datasets, domain(s), model, and performance parameters (precision, recall, F-1, or accuracy). This study

Table 7
Aspect Term Extraction (ATE) using CNN, RNN, and its variants models.

Paper	Domain	Dataset	Model	Performance		
				Precision (%)	Recall (%)	F1 (%)
[73]	5 products Electronic	[1]	CNN + Amazon embedding + POS + Linguistic patterns	92.75	88.32	90.44
	RES	SE-14		88.27	86.1	87.1
	LAP	SE-14		86.72	78.35	82.32
[58]	RES	SE-16	RNN + Google WE	75.49	69.44	72.34
[72]	RES	SE-16	Bi-LSTM + Google WE + CRF			72.44
[71]	RES	SE-14	LSTM + POS + Chunk + Amazon embedding			82.06
	LAP	SE-14				75
[120]	RES	SE-14	Holo DyMEMNN	81.87	79.73	79.73
	LAP	SE-14		75.16	73.19	74.03
[121]	RES	SE-14	Bi-LSTM + Senna embedding + Local Context			80.62
	LAP	SE-14				74.78
[122]	RES	SE-14	Uni-directional Elman RNN			82.12
	LAP	SE-14				75.45
[88]	RES	SE-14	ReNN + CRF + POS + Sentiment Lexicon			84.9
	LAP	SE-14				78.4
[123]	RES	SE-14	Aligning aspect embedding			79.10
	LAP	SE-16				82.89
	RES	SE-16				74.09
[86]	RES	SE-16	Hierarchical bidirectional LSTM			85.3
	LAP	SE-16				80.1
[124]	RES	SE-14	Gated Alternate Neural Network			80.09
	LAP	SE-14				72.21
[125]	RES	SE-14	Lexicon-based + corpus-based	91	91	91
[126]	LAP	SE-14	Customizations of LSTM + word/character embeddings			81.08
	RES	SE-14				86.05
[127]	LAP	SE-14	Unified position-aware CNN	Accuracy: 73.73		
	RES			Accuracy: 80.05		
[128]	LAP	SE-14	Hierarchical Multi-task Learning			80.83
	RES					85.12
[129]	RES	SE-14 SE-15	RNN+ Auxiliary labels + Hierarchical network (Cross-domain)			77.9
	LAP	SE-15				76.6
[130]	LAP	SE-14	Memory interaction network with LSTM + Extended memory			77.58
	RES	SE-16				73.44
[85]	LAP RES	Challenge ESWC 2016	Bi-GRU + Amazon WE + POS	65.9	71	68.4
[131]	RES	SE-14	Multitask learning NN combines Bi-LSTM and CNN layers			83.65
		SE-15				67.73
		SE-16				72.95
[132]	RES	SE-14	CNN + Dependency-Tree + POS			83.97
	LAP	SE-14				75.66

* Restaurant: RES, Laptop: LAP, SemEval-14: SE-14, SemEval-15: SE-15, SemEval-16: SE-16.

only considers English reviews for both the ATE (Tables 7 to 10) and ACD (Table 11).

8.1. Aspect term extraction (CNN and RNN based deep learning model)

The idea behind the contribution of the CNN model to ABSA tasks is that the majority of the aspect terms are either nouns or phrases of nouns. The CNN model is used in ABSA tasks because it is believed that, irrespective of the order in which they appear, key words may contain the aspect term, denote a category, or establish polarity. With its architecture, the CNN is able to develop the ability to find certain features. Besides, ATE, ACD and sentiment aggregation do not depend on the position of keywords in the sentences. Goldberg [100] suggested that these position-independent features be utilized for extracting fixed-length nouns/noun phrases. Further, Poria et al. [73] promoted CNN models as non-linear in learning; hence, they do not require manually generated features (language rules). As compared to linear models (CRF, etc.), CNN better represents the data. Liu [87] concatenated part-of-speech (POS) with word embedding and provided this information as a training resource for the DNN model. Wu et al. [61] applied a GRU-based hybrid technique that combines rules and a supervised approach for aspect extraction tasks. Table 7 and Table 11 show the articles in which CNN has applied for the ATE and ACD subtasks of ABSA, respectively.

In the tables presented in this section, SemEval-14 is represented as SE-14, whereas SemEval-15 and SemEval-16 are represented as SE-15 and SE-16, respectively. On the other side, Restaurant and Laptop are represented as *RES* and *LAP*, respectively. Poria et al. [73] have successfully used the CNN-based architecture combined with linguistic patterns for word-level aspect extraction on product reviews, given by Hu and Liu [1] and the SemEval-14 datasets. They formed a five-word feature window for each word in the sentence, considering that the neighboring words contribute to the individual word's tagging. Research findings indicate that for ATE subtask, the deep CNN model outperformed conventional DNN models and attained remarkable accuracy. Toh and Su [58] have used the RNN model to achieve the best possible outcome with Google embedding on the SemEval-16 dataset. RNN models with CRF as a classification layer have significantly contributed in ATE subtask for understanding the contextual meaning of the review by capturing long-term and inter-sentence dependencies. Chen et al. [72] attained cutting-edge findings by integrating the outcomes of Bi-LSTM and CRF (for IOB tagging) on the restaurant domain of SemEval-16. This model captured long-term dependencies using contextual information and improved the fixed window size issues of CNN models. According to Liu et al. [71], LSTM combined with linguistic patterns of POS, Amazon word embedding, and word chunks achieved much better results than CRF as a classification layer.

Table 8
Hybrid deep learning domain-specific aspect extraction approaches.

Paper	Domain	Dataset	Model	Performance		
				Precision (%)	Recall (%)	F1 (%)
[133]	RES	SE-14	Double embedding + Bi-LSTM + Triplet encoding			63.53
	LAP	SE-14				50.0
	RES	SE-15				54.21
	RES	SE-16				63.21
[134]	RES	SE-14	Position less + Bi-GRU	Accuracy: 81.37		
	LAP	SE-14		Accuracy: 75.39		
	RES	SE-15		Accuracy: 80.88		
	RES	SE-16		Accuracy: 89.30		
[18]	RES	SE-16	Hybrid + Rule-based + Attention	81.02	77.09	79.01
	LAP	SE-16		63.23	74.57	68.43
[61]	RES	SE-15	Bi-GRU + WE + POS + unsupervised linguistics	62.26	65.13	63.36
	RES	SE-16		58.94	70.59	64.24
	RES	SE-14		72.17	73.43	72.8
	LAP	SE-14		64.29	74.02	68.81
[106]	Cell phone	Twitter	Deep CNN + Rule-based method + Clustering	79.24	88.4	82.3
	Camera			78.79	89.9	80.5
	LAP	SE-14		79.25	88.45	83.2
	RES			79.67	86.2	83.3
	Movie data	Twitter		78.67	82.2	80.3
[135]	RES	SE-16	Cross-domain + Hybrid + Attention	63.42	74.20	
	LAP			58.68	78.38	
[118]	Book	Amazon Product reviews	Domain attention model + multi domain	Accuracy: 87.75		
	DVD			Accuracy: 86.58		
	Electronics			Accuracy: 87.70		
	Kitchen			Accuracy: 88.93		
	Apple	Sanders Reviews		Accuracy: 86.76		
	Google			Accuracy: 89.46		
	Microsoft			Accuracy: 86.36		
	Twitter			Accuracy: 82.71		
[136]	RES	SE-16	Hybrid + Attention model + Domain ontology	Accuracy: 87.1		
[137]	20 domains	Dranziera	Neural word embedding + multi-domain	85.15	84.07	84.6
	8 domains	Non-Dranziera		84.42	83.43	83.9

* Restaurant: RES, Laptop: LAP, SemEval-14: SE-14, SemEval-15: SE-15, SemEval-16: SE-16.

Recently, Kajdanowicz and Kazienko [126] improved the performance using customizations of LSTM and word/character embeddings. In addition, a comprehensive investigation of varieties of LSTM-based models and embedding vectors is presented. Li and Lam [130] proposed a Memory Interaction Network with LSTM, whereas Liu and Shen [124] improved the existing work with the Gate Truncation RNN (GTR), in which convolution and pooling are applied to obtain the correlative space among domain-specific words and aspect terms. Two more lexicon generating techniques are proposed by Mowlaei et al. [125] or the ATE subtask, one applying statistical approaches and the other one utilizing a genetic algorithm. This model achieved prominent results with lexicon-based and corpus-based aspect extraction approaches. Wang et al. [97] proposed a unified position-aware CNN model and further improved the model with hierarchical multi-task learning [127]. Although the above-discussed RNNs with CRF enhanced aspect extraction efficiency, these approaches cannot utilize past and future contexts. Ding et al. [129] have used the bidirectional RNN model, taking into account the context of each word in the lengthy statement, both present and past. Further, Jebbara and Cimiano [85] have extended the aspect extraction performance and combined Bi-GRU with pre-trained semantic Amazon embedding, Word-Net, SenticNet, and POS. The effectiveness of bi-RNN models (Bi-LSTM, Bi-GRU) is considered as an alternative to the topic. By methodically revisiting the obstructions on all conceivable sides, comprised of the model, data, and training, Fei et al. [138] increased the ABSA's robustness. According to research, considering relationships among aspects and context as attributes, analyze such aspect-invariant opinion statements. Out of the blue can cause noise to be introduced, especially if there are not enough aspect-related annotations [139]. Using multi-label learning techniques, Kumar et al. [140] developed a gender-specific multi-aspect sentiment identification framework for Medicare applications.

Next, the above-discussed research focused on a multitask learning approach that handles the sentiment analysis issue as a table-filling problem while simultaneously extracting aspects and sentiment terms. The issue of recognizing overlapped sentiment triples is resolved by the multitask learning structure, but the entire framework may directly replicate connections among aspects and opinions. Dai et al. [133] present a dual-table design used to support a sentiment-dependence detector that begins with two angles: aspect to opinion and opinion to aspect, to construct two tables of sentiment-dependence that are predominately composed of two different sorts of data.

In recent times, researchers have started contributing to another promising application for aspect extraction: combining CNN and RNN-based models to utilize the importance of rule-based linguistic patterns with DNNs [102,132]. Table 8 presents some hybrid deep learning approaches for aspect extraction. These approaches combine unsupervised methods with a DNN to improve the ATE and ACD tasks of ABSA. First, the nouns and noun phrases are captured as aspect words using linguistic norms by resolving the language constraints. Next, the retrieved attributes are utilized for modeling a Bi-LSTM model in a deep learning network in the first stage as labeled data for better prediction of aspects. Ray and Chakrabarti [106] used hybrid deep CNN and a linguistic method to enhance aspect extraction performance on product reviews from Twitter and SemEval-14 datasets. The domain-specific aspect extraction approach is further extended using an NP-chunk-based hybrid unsupervised method with the Bi-GRU network [61] and a hybrid attention-based model [24], presented in Fig. 15.

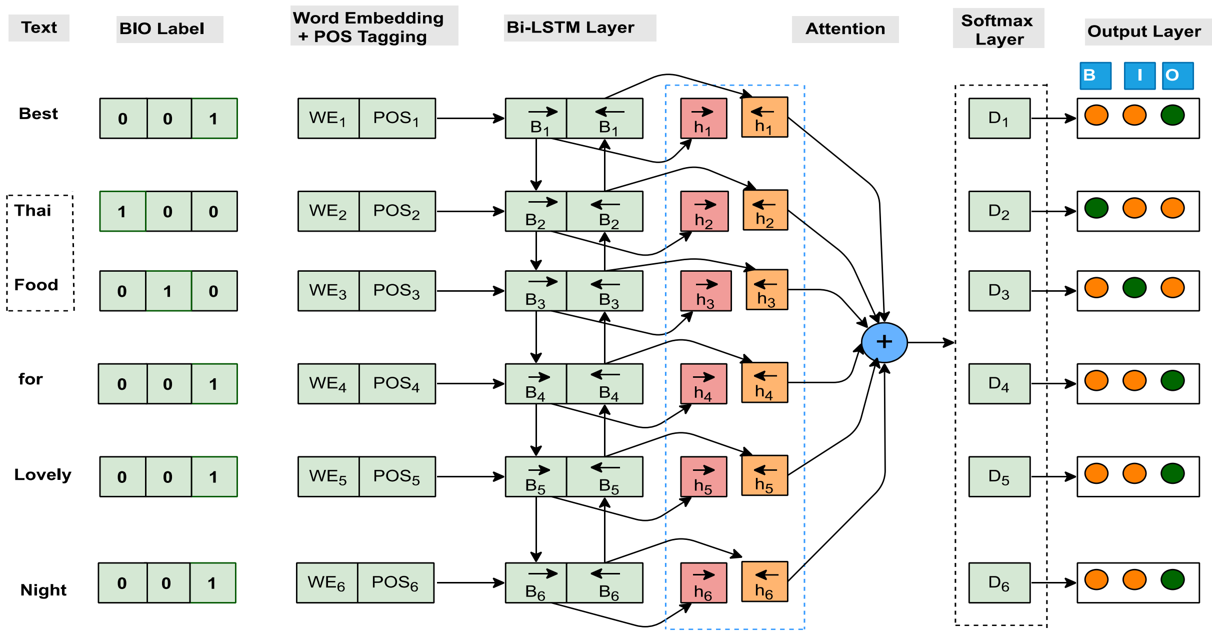

Fig. 15. Hybrid Attention + Rule-based method for ABSA from [24].

Ding et al. [129] have further provided an unsupervised rule-based method for auxiliary sequence labels of review and then used a recurrent neural network for aspect extraction for multiple domains. These hybrid approaches are beneficial for the domains, which bears no labels, so even some of the best-performing supervised deep learning approaches will not be effective on it. Applying current algorithms to a new domain or language is typically time-consuming and costly due to the fact that they typically need a substantial quantity of labeled data or outside linguistic assets. By utilizing only a few seed phrases for every component category and sentiments, Zhuang et al. [141] hope to develop an unlabeled corpus' aspect-based sentiment analysis model with the least amount of user input possible. In further advancement in the research, Dragoni and Petrucci [137] proposed neural word embedding for multi-domains; addressing multi-domain Amazon and Sander reviews, a domain attention approach has been presented in [118], which can extract the most prominent features from the hidden layer. When a model is put into a domain that is different from the domain on which it was trained, the existing methodologies struggle with domain adaptation, which results in inadequate results. To accurately train a model and extract the most pertinent domain-specific elements, you need a sizable sample corpus for each domain [62].

As indicated in Tables 7 and 8, the DNN models explained have achieved significant results on Semeval-14; still, performance is not promising on Semeval-16. The fact that the dataset includes evaluations with several sentences is one of the main causes, and previous approaches failed to establish inter-sentence dependencies. Applying pre-trained aspect aligned embedding, Tan et al. [123] increased aspect extraction precision although only became able to extract aspects at the sentence-level on individual sentence evaluations. As shown in Fig. 16, taken from [86], a hierarchical bi-LSTM model on five multilingual and multi-domain domains for modeling inter-sentence as well as intra-sentence dependencies is presented. It is still difficult to incorporate conceptual data to link ambient and feature terms. Hence, Chauhan et al. [19] intends to include a sentence coreference resolution phase before executing the BERT-based hierarchical attention method for the ABSA system to eliminate insufficient information and capture aspect referencing.

Using ABSA in recommendation systems has also been an encouraging trend in recent years. Musto et al. [142] offer a method for automatically extracting pertinent and distinctive features of the suggested product based on customer reviews in order to support the proposals made by a recommendation system. Regarding rating prediction challenges, Liu et al. [94] suggest a unique Multilingual Review-aware Deep Recommendation Model (MrRec) based on ABSA, which is further extended by Pastore et al. [143]. Positional embedding was used in several previous studies to highlight the connection between an aspect word and its surroundings, and the results showed good ABSA accuracy. Most often, the disparity between the aspect word and the rest of the contextual words, also referred to as the position indicator sequence, determines the positioning encoding. To get the most cutting-edge achievement, such strategies typically combine complicated preprocessing methods with extra-attainable positional embedding using complex architectures. Yadav et al. [134] streamline preprocessing by incorporating polarity lexicon substitution and filtering approaches, which convey the positional characteristics of the aspect word and do away with positional embedding.

8.2. Impact of attention mechanism on aspect term extraction

It is observed in recent research that the aspect extraction task is not evenly impacted by all the terms in the review sentence. As explained in Section 7.1, the mechanism of attention is yet another powerful strategy to encapsulate the prominent impact of the words on the review's sentiment. Wang et al. [114] have applied GRU-based multilayer coupled attention for extracting aspects and opinions altogether, depicted in Fig. 17. Akhtar et al. [144] offered a multitask learning system for aspect extraction and category detection includes a self-attention mechanism and a Bi-LSTM.

Additionally, the attention mechanism is included to resolve the domain adaptation issues. Extraction of cross-domain aspect terms and classification of aspect-level sentiment using an integrated framework are presented in [146]. A hybrid attention-based approach given by Singh Chauhan et al. [18] has improved the aspect extraction accuracy for restaurant and laptop domains. Further, a residual attention mechanism [115] has been put up to

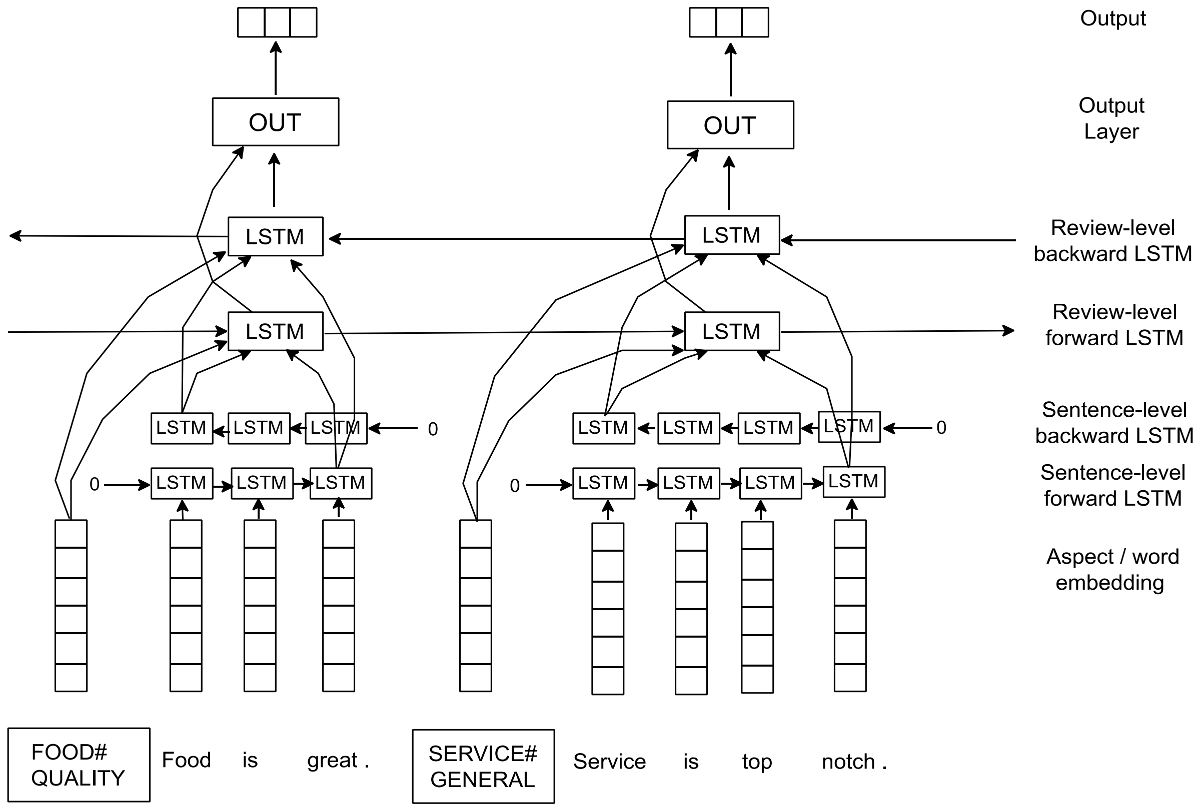

Fig. 16. Hierarchical Bi-LSTM model for Co-referencing multi-sentences review [86].

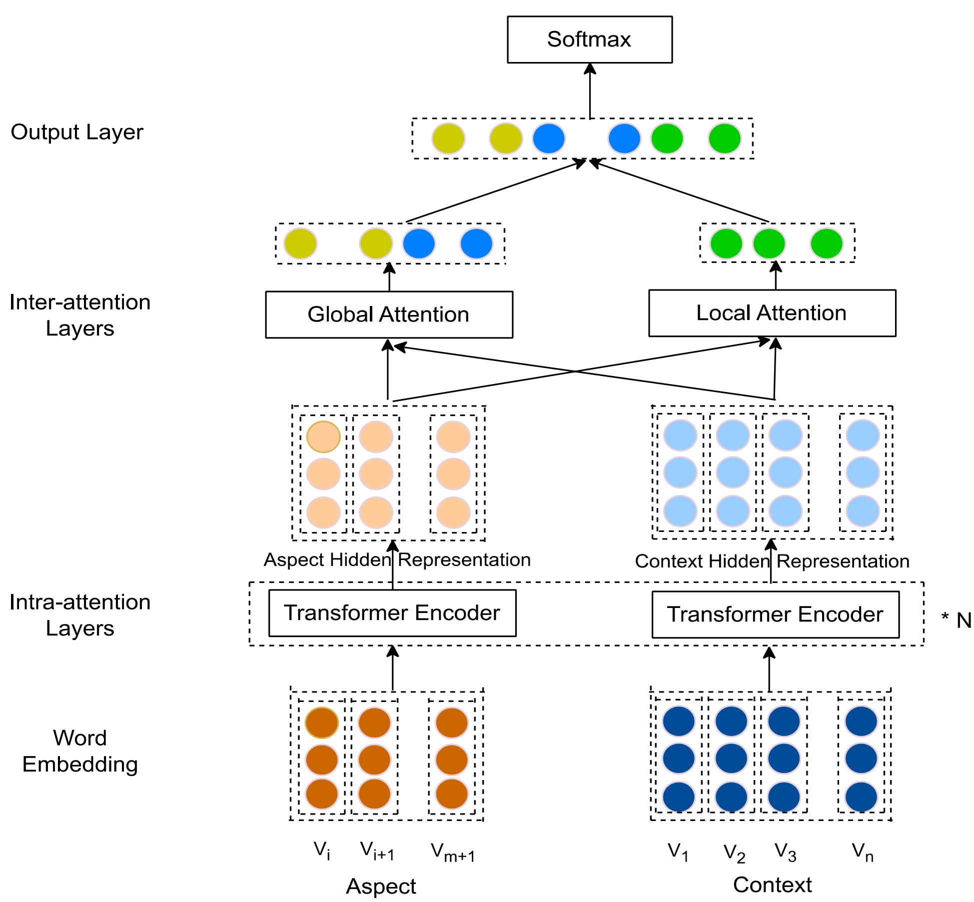

Fig. 17. Aspect Extraction using multi (inter and intra) layer attention [145].

Table 9
Aspect extraction using attention-based mechanisms.

Paper	Domain	Dataset	Model	Performance		
				Precision (%)	Recall (%)	F1 (%)
[144]	RES	SE-14	Bi-LSTM + Self-Attention	82.83	80.43	81.61
	LAP	SE-14		75.40	63.30	57.26
[149]	RES	SE-14	Context and Memory	Accuracy: 82.57		
	LAP	SE-14	Network+ Self-Attention	Accuracy: 77.46		
[148]	LAP	SE-14	Truncated History-Attention			79.52
	RES	SE-14	and Selective Transformation			85.61
	RES	SE-15	built on two LSTMs			71.46
	RES	SE-16				73.61
[18]	RES	SE-16	Hybrid + Attention	81.02	77.09	79.01
	LAP	SE-16		63.23	74.57	68.43
[115])	RES	SE-16	Residual attention			71.34
	RES	SE-16				69.72
[122]	RES	SE-14	Coupled multi-layer			85.29
	RES	SE-15	attention-based on GRU			70.33
	LAP	SE-14				77.8
[150]	LAP	SE-14	Filter gate network +	Accuracy: 81.37		
	RES	SE-14	Multi-Head Attention	Accuracy: 83.94		
	RES	SE-16		Accuracy: 85.93		
[116]	RES	SE-14	Co-Attention LSTM/			78.8
			Co-Attention MemNet			79.7
	LAP	SE-14				73.5
						72.9
	Twitter	[21]				71.8
						70.8
[151]	RES	SE-14	Lightweight model DNet +	Accuracy: 79.38		
	LAP	SE-14	gated CNN+ Attention	Accuracy: 73.98		
[118]	Book	Amazon	Domain Attention model +	Accuracy: 87.75		
	DVD	Product	Multi Domain	Accuracy: 86.58		
	Electronics	reviews		Accuracy: 87.70		
	Kitchen			Accuracy: 88.93		
	Apple	Sanders		Accuracy: 86.76		
	Google	Reviews		Accuracy: 89.46		
	Microsoft			Accuracy: 86.36		
	Twitter			Accuracy: 82.71		
[119]	RES	SE-14	Augmented Knowledge	Accuracy: 75.87		
			Sequence-Attention Mechanism	Accuracy: 72.42		
	LAP	SE-14				
[145]	LAP	SE-14	Multi-Attention Network	Accuracy: 78.13		
	RES	SE-14		Accuracy: 84.38		
	RES	SE-16		Accuracy: 85.87		
[147]	RES	SE-14	Context Aware Attention Network	Accuracy: 86.25		
		SE-16		Accuracy: 89.56		
[152]	LAP	SE-14	Content Attention Model	Accuracy: 75.07		
	RES			Accuracy: 80.89		
[153]	LAP	SE-14	Component focusing multi-head	Accuracy: 79.44		
	RES		co-attention networks	Accuracy: 84.48		
	Twitter			Accuracy: 74.22		

determine the semantic relationship between every component word and its linguistic scenario. Several other attention-driven RNN networks have effectively obtained remarkable results on restaurant and laptop evaluations during the past couple of years, such as an alternate co-attention framework for target-level and scenario-level learning for efficient context representation [116], a neural context-aware neural attention network [147], and a self-directed iterative attention-based aspect extraction method [96,148] used Truncated History-Attention (THA) and Selective Transformation Network (STN); Yang et al. [116] used Co-attention LSTM networks and Co-attention MemNet networks as presented in Table 9. An improved embedding in the token words and language constraints with filtered aspect extraction for the convolutional network using attention mechanisms has improved the aspect extraction performance [119].

Generally, existing methods require extra training time due to sequence models. Xu et al. [145] utilized both intra as well as inter-level attention in which, first, in addition to maintaining long-distance opinion relationships, a transformer encoder concurrently encapsulates the input review. Next, global and local attention are used to associate the aspect with context information. Global attention captures the entire relationship; on the other side, local attention maintains correspondence at the word level. The existing methods usually ignore local and global attention mechanisms. The performance is further enhanced with a brand-new attention network (SDATT) that explicitly takes into account grammatical dependencies [154]. The context and aspect memory network (CAMN) were applied a multi-attention model combined with a bi-LSTM. The first stage uses self-attention to capture the context information, the second stage, although uses encoder–decoder attention to associate the context with the aspect [149]. Considering the several head's attention (FGNMH) approach, the Filter Gate Network first uses the concatenated context corpus with part-of-speech tagging as input representation. It then uses multi-head attention to model domain-specific semantic information. In the end, a filter layer prunes out the domain irrelevant aspect words [150].

Liu et al. [155] incorporated a module for dual-feature retrieval to acquire aspects-related and aspects-unrelated aspects utilizing graph convolutional networks and attention methods. Liu et al. [152] put out a novel content-attention-based ABSA framework that incorporates two strategies for improving attention:

the context attention model and sentence-level content attention mechanisms. Despite the remarkable improvement, there are still several difficulties with ABSA in real society. (1) The existing attention-based approach may lead a given aspect to focus attention on incorrect phrases which have no relationship syntactically. (2) The sentiment expressed by unique sentence structures, such as double negatives, is not captured by conventional approaches. (3) The majority of research uses a single vector to represent the background and the target [156].

In order to address the above-discussed issues, Cheng et al. [153] suggest a component-centered multi-headed co-attention network framework that, in order to learn target representation more precisely, devotes focus to the target pattern through targeted expanded context. To record the connections across the surroundings and a specific component, attention methods are frequently used. The classification engine's ultimate representation is the weighted sum of context-concealed states. Xing and Tsang [157] create a three-channel set of aspect-aware contextual encoders using an aspect-aware LSTM, an aspect-aware GCN, and aspect-aware BERTs. Even though new research favors using self-attention networks to complete the ABSA challenge, deriving long-distance relationships among non-adjacent words remains challenging, particularly if a paragraph has multiple aspects. Mokhosi et al. [158] suggest the BERT-MAM approach, which sees the ABSA issue as a process of memory activation caused by word similarity and memory deterioration. This model assumes that a word's significance gradually declines until it is reactivated by a similarity increase. A novel method for ABSA further improved the performance, which divided the problem into three distinct sequential labeling parts: target tagging, opinion tagging, and sentiment tagging [159].

8.3. Impact of BERT on aspect term extraction

Different attention-driven deep learning approaches have been developed during the past few years to capture the context of different aspect terms within the review sentences. As a result, the attention mechanism has remarkably contributed to ABSA. Even so, the attention model is only capable of lengthening the training period for long review sentences by providing extra weight to the model's attributes. The lack of training information for jobs that are related to particular domains is one of ABSA's largest problems.

ATE is a complex process with several domains that calls for diverse contextualize word vectors for an identical word in accordance with various domains. In the BERT-based technique, first pre-train the general-purpose language representation model of a neural network using a huge amount of un-annotated data on a known domain. Next, The trained model was used as the foundation to fine-tune an entirely novel job-specific framework [75,91]. In RNN and its variants, backpropagation, vanishing gradient, and exploration gradient issues frequently appear in model training. The accuracy of the previous work mainly relies on static word embedding, which produces comparable word representations for an identical term across domains. It is still very difficult to incorporate conceptual information for linking the context with the aspect terms in a multi-sentence evaluation. By combining opinion domain understanding alongside the input's word representation, the BERT model combines the extra information from a sentiment knowledge graph. It also improves performance even with sparse training data by incorporating external domain knowledge into the language representation model.

Further, aspect extraction approaches semantically associate target and context words based on RNNs or pre-trained models such as BERT. As explained in Section 6.2, BERT is intended to collectively condition both left and right context in all layers in order to pre-train deep bidirectional representations from the unlabeled text. As a result, the pre-trained BERT model only requires a single extra output layer for fine-tuning to produce cutting-edge models for a variety of jobs. The investigational findings of the study, which are shown in Table 10, demonstrate that BERT trains the context in a domain-particular corpora and incorporates the POS elements of the context to improve the representation of the context as a whole. BERT-based input representation is also helpful to increase the precision of the aspect extraction process from multiple sentences of the review. Furthermore, for the extraction of contextual aspects, Zhao and Yu [92] suggest the knowledge-enabled linguistic representations BERT. Integrating opinion domain expertise with the input word representation, this framework incorporates the extra information from an opinion information graph, which combines adjacent domain understanding into the linguistic presentation model to further improve performance despite sparse training data. Generally, aspect extraction approaches semantically associate target and context words based on RNNs or pre-trained models such as BERT. In the recent works of ABSA, Li et al. [93] have presented a BERT-based aspect extraction architecture and outperformed baseline models on restaurant and laptop data of SemEval-16, as shown in Fig. 18. Sun et al. [95] have utilized BERT for aspect extraction with auxiliary sentence construction, and Hoang and Rouces [164] show the potential of contextual word representations with BERT to solve cross-domain aspect extraction problems.

These days, researchers primarily concentrate on an individual job, or a combined job made up of the various jobs; yet a single job or joint job study is insufficient to fully understand sentiment analysis. For this combined task, Zhang et al. [34] improved the accuracy by introducing a neural network model with two stages made up of a number of components, for instance, Bi-LSTM, straightforward gated self-attention, and location encoding. Considering the popularity and effectiveness of the procedures, the current GCN-based models lack an effective restriction technique for the communication transmission to the aspect terms, which results in significant noise during graph convolution. Furthermore, they do not fully utilize the capabilities of BERT and just aggregate using the BERT sub-word vectors to produce word embeddings. To solve these issues, Zhao et al. [165] propose a graph convolutional network with several weight methods for aspect-based sentiment present a graph convolutional network with multiple weight approaches for ABSA.

It has been observed and investigated that the discussed aspect extraction methods degrade performance as they cannot correctly adjust standard lexicons to the dataset context. This issue is considered using the BERT Base Uncased model, a powerful Deep Learning Model [162], which achieved comparable results. On the other hand, BERT-based multiple attention mechanisms improve domain-specific aspect term extraction [97]. A BERT-based contextual co-attention network further improves performance [94]. In order to detect personalities, Ren et al. [166] used BERT-driven phrase-level meaning retrieval while taking opinion information into account. In [161], a new Syntax-and Knowledge-based Graph Convolutional Network (SK-GCN) aspect word extraction in the model using BERT makes advantage of the syntax dependent tree and natural instincts via GCN. In recent times, several ways have been developed to make use of contrastive methods of learning to improve ABSA's accuracy through the acquisition of aspect-based emotion representations. Xu and Wang [167], unsupervised contrastive learning based on augmentation and sentiment-based supervised contrastive learning strategies for improving ABSA results are presented and contrasted. While neglecting the sentiment links between various components, the majority of currently used methods treat various parts of a sentence separately.

Table 10
BERT-based deep learning models for aspect extraction.

Paper	Domain	Dataset	Model	Performance		
				Precision (%)	Recall (%)	F1 (%)
[160]	Car Phone Notebook Camera	Product	BERT + BILSTM + Self-Attention	F1: 94.71 F1: 94.69 F1: 93.59 F1: 93.65		
[158]	LAP RES Twitter	SE-14 Twitter	Maximal activation weighted memory + BERT	Accuracy: 83.10 Accuracy: 87.10 Accuracy: 75.80		
[19]	LAP RES	SE-16 SE-16	BERT + Attention using Co-reference resolution	68.08 85.18	75.78 81.83	71.81 83.20
[154]	LAP RES RES	SE-14 SE-14 SE-16	BERT + Syntactic dependency + Attention Network	Accuracy: 81.89 Accuracy: 88.97 Accuracy: 92.32		
[93]	RES LAP	SE-16 SE-16	BERT representation + GRU + Self-Attention	70.61 61.88	76.20 60.47	73.2 61.1
[161]	LAP RES RES	SE-14 SE-14 SE-16	Graph Convolutional Network-SK-GCN + BERT	Accuracy: 79.00 Accuracy: 83.48 Accuracy: 87.19		
[128]	LAP RES RES	SE-14 SE-14 SE-16	Multiple Attention + BERT	Accuracy: 81.35 Accuracy: 86.52 Accuracy: 90.34		
[162]	Product	Amazon	Fine-tuned BERT Base Uncased model	88.09	86.22	89.41
[96]	RES LAP	SE-141 SE-14	Self-Attention in BERT			87.86 82.64
[117]	RES LAP Twitter	SE-14	BERT-based Co-Attention Network			81.15 76.22 75.61
[163]	LAP RES Twitter	SE-14 SE-14	BERT + Target-dependent Network	Accuracy: 78.87 Accuracy: 85.27 Accuracy: 77.31		

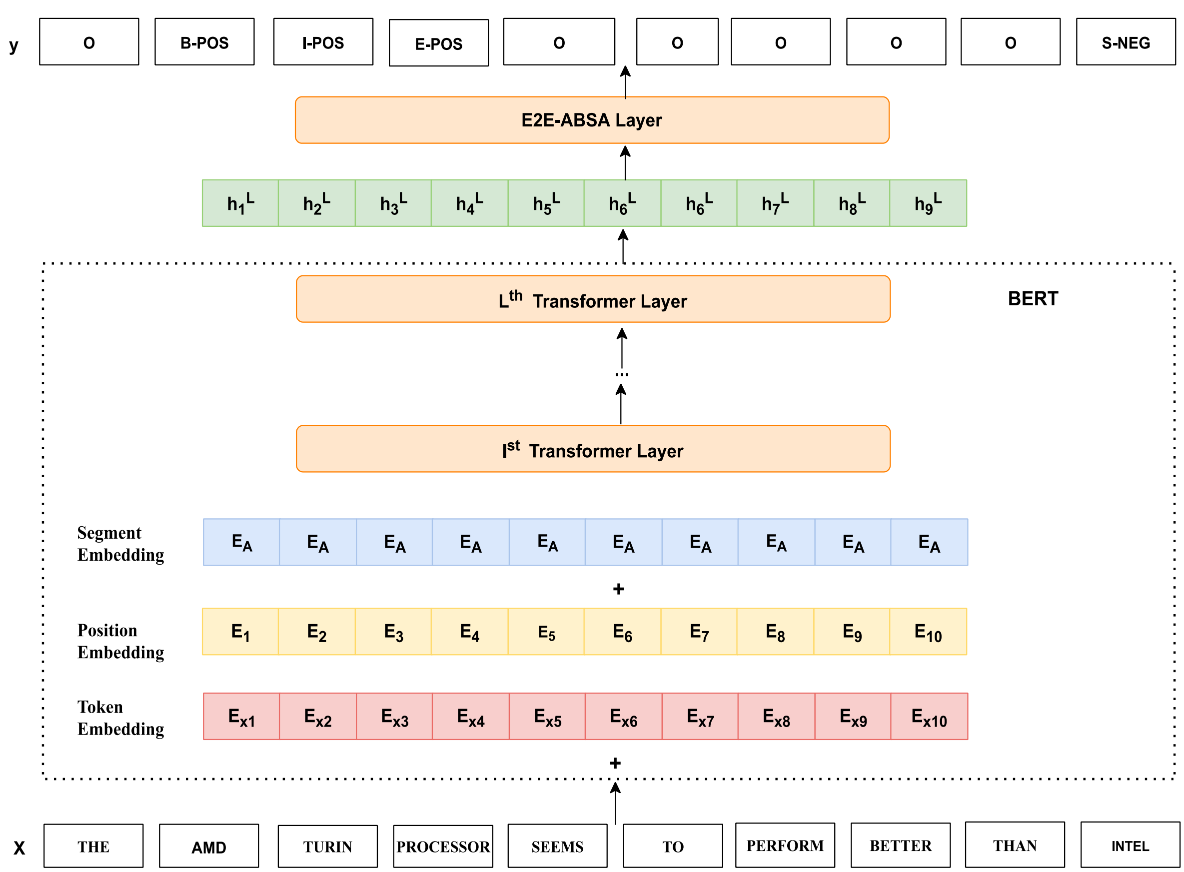

Fig. 18. DNN model with BERT embedding Layer for Aspect Extraction.

Table 11
Aspect Category Detection (ACD) using CNN, RNN and its variants models.

Paper	Domain	Dataset	Model	Performance		
				Precision (%)	Recall (%)	F1 (%)
[58]	RES	SE-16	CNN + Word embedding +			75.1
	LAP	SE-16	Name list + Head word			59.83
[105]	Smartphone	Amazon	Multi-CNN for each aspect			83.74
[171]	Smartphone	Amazon	Multi-task CNN + Word embedding			81.2
[120]	RES	SE-14	Tensor DyMEMNN			81.68
	Reviews	SE-14 SE-15				81.66
[129]	RES	Yelp dataset	RNN + Word Embedding + finetuning			72.42
	LAP	Amazon product review				66.1
[172]	Hotel	DBS challenge 2015	Weighted hierarchical Bi-RNN			65
[88]	RES	SE-14	Coupled Multi-layer Attention based on GRU			85.29
	RES	SE-15				70.33
	LAP	SE-14				77.8
[148]	LAP	SE-14	Memory Interaction Network (MIN)based on LSTM with extended memory			77.58
	RES	SE-16				73.44
[131]	RES	SE-14	Multitask learning NN combines Bi-LSTM and CNN layers			88.91
		SE-15				65.97
		SE-16				76.42
[173]	RES	SE-14	Association rule mining+ Co-occurrence frequency Data + unsupervised method	67.5	64.7	67
[99]	Reviews	Updated SE-15	LSTM Target Attention + SenticNet			76.44
[174]	RES	SE-16	W2VLDA + Topic Modeling + multi-domain	70	77	72
[106]	Cell Phone	Twitter	Deep CNN + Rule-based method + Hierarchical clustering	79.24	88.4	82.34
	Camera			78.79	89.9	80.5
	LAP	SE-14		79.25	88.45	83.24
	RES			79.67	86.2	83.34
	Movie Review	Twitter		78.67	82.2	80.34

Further, to represent the latent correlations across various elements in the emotional subspace, Li et al. [168] present an Aspect-Pair Supervised Contrastive Learning (APSCL) model. In addition, Zhao et al. [169] use multi-head attention (MHA) to link interdependence patterns together in aspect gathering, combining the two separate ATE objectives and emphasizing the important dependency linkages, allowing the system to concentrate on the phrases that have a strong connection to the aspect. The modeling of phrase syntactic structures using graph-oriented techniques, such as GCNs, graph attention networks (GATs), with its expansions, has recently shown encouraging results. Such graph-based techniques, however, typically homogenize relationships and lose semantic knowledge. Additionally, systems that simply employ syntax-based graph techniques fail to recognize the emotions present in unstructured comments. To enhance the effectiveness of aspect extraction, Zhou et al. [170] advocate using two specifically designed GCNs for collecting linguistic and semantic details, called the dependency feature-aware hierarchical dual-graph convolutional network (HD-GCN) model.

The discussion that was just had shed light on the various strategies that researchers have used for NLP tasks. It is obvious that outcomes depend on the semantics being investigated as well as the model selection. The selection of a model was dependent on the overall semantics of a classification task, hence hyper-parameters like layer size ought to constitute the main consideration.

8.4. Aspect category detection

Table 11 shows that CNN-based models have contributed more to the ABSA subtask than ATE, such as ACD. Toh and Su [58] have combined CNN with word clusters, name lists, and headwords. They used binary relevance approaches to categorize the aspects into predefined aspect categories and assumed that the CNN features contributed maximum performance. Wu et al. [171] proposed multi-task CNN with word2vec on Amazon product reviews. Gu et al. [105] further improved the performance of Multi-task CNN to overcome the burden of feature engineering by cascading multiple CNN models (aspect mappers and sentiment classifiers) of each aspect on Amazon product reviews.

Next, the RNN based model with bidirectional variations is also successfully applied to identify aspect categories, such as [129] using hierarchical LSTM. Chaudhuri and Ghosh [172] have applied weighted hierarchical Bi-RNN from the DBS text mining challenge on to hotel reviews. They have used the model with four bi-RNN layers, two other layers: a softmax layer and one fully linked layer. The attention mechanism has also been applied for the ACD task in which, aspect embedding has been used for deciding attention weights for aspect categorization [99,114]. Fig. 17 shows that attention concentrates on different important parts for different aspects. A memory network, Tensor DyMemNN, models the dyadic interaction between different words and nouns and noun phrases in the sentence [120,130] proposed a Memory Interaction Network with LSTM, whereas [122] applied GRU-based multilayer coupled attention for ACD. Xue et al. [131] have proposed a hybrid approach and assumed that aspect extraction and aspect category detection are closely related subtasks of ABSA, in which bi-LSTM (ATE) and CNN (ACD) mutually share information with a multi-task learning neural network. García-Pablos et al. [174] proposed a topic modeling-based, almost unsupervised system called W2VLDA for multi-domain aspect category detection. Ray and Chakrabarti [106] have used a hierarchical clustering approach for categorizing aspects in a pre-defined set of categories.

8.5. Conclusion remarks of using deep learning for aspect extraction

The above-detailed discussion has provided the direction of different deep learning approaches used by researchers for various subtasks of ABSA. The research of numerous methodologies has demonstrated that the type regarding information that is evaluated has a significant impact on the effectiveness of systems. Besides, in model training, the model's hyper-parameters (layer size) played an important part in ATE and ACD subtasks. Further, it was observed that CNN could extract the most important review patterns [98]; hence, the model could produce fixed-length input for hidden layers [73]. On the other hand, the above discussion and the study of Xu et al. [102] have shown that CNN-based approaches require large training data as well as parameters needed for fine-tuning the system. Further, it was also observed that CNN models failed to parse long sentences due to the fixed size of the hidden layer; hence, it is not feasible to identify enduring dependencies and the contextual meaning of the long reviews [88]. Some CNN-based approaches, such as [73], propose to form feature windows of fixed size by combining neighboring words of each word; still, it is impossible to extract the context information outside the feature window. Furthermore, it is also not possible to establish a correlation between multi-sentence reviews using CNN models. In summary, CNN-based approaches are useful for short sentences and contain one target or opinion of fixed length.

Having the ability to capture long-term dependencies, RNNs and its variations of long sentences and multi-sentence reviews have greatly enhanced ABSA's performance on benchmark datasets. The main contribution of RNNs is that they consist of distribution hidden states, which make them capable of storing memory and fitting non-linear data; they do not require a large set of training data. These properties make RNN more suitable for ABSA subtasks than CNN if aspect extraction requires contextual information from the complete review. However, the RNN models are not useful where the sentences are represented using a key phrase.

Tables 7 to 11 observe that most of the studies for ATE and ACD used review data in English from restaurant and laptop domains of SemEval-14 and SemEval-16. The performance of DNNs on SemEval-14 is better than SemEval-16. RNN and its various versions are used as the foundation for a significant portion of ATE techniques. The main variants of LSTM/GRU using bidirectional RNNs are hybrid unsupervised rule-based methods, attention mechanisms, cross-domain, and hierarchical approaches, etc. The studies evidently showed lower performance for ACD and ATE subtasks on reviews of laptops in contrast to restaurant reviews because of the higher number of noun phrase aspects. Further, the accuracy value of the restaurant reviews is superior to the laptop reviews because the laptop reviews hold fewer contextual aspects. In addition, laptop reviews recall is lower because they hold a variety of multi-word aspects.

Due to the lack of one word as well as noun phrases as features in the laptop domain, the entire F-score of aspect extraction is smaller in comparison to that of the restaurant reviews. Several sentences exist in the SemEval-16 dataset where a single aspect contains multiple aspect categories in distinct sections of the sentence. The laptop domain contains more varieties (81) of aspect categories than the restaurant domain (12). Moreover, the aspect may appear in distinct sections of the long sentences; therefore, the Bi-LSTM may miss several relevant aspects and detect the wrong aspect category. Further, a review containing multiple sentences holds inter-sentence and intra-sentence dependency, which, as a result, identifies wrong aspects and aspect words for improper nouns and noun phrases. Next, the aspect term extraction needs to use interdependency between sentences to identify the proper aspect category.

In addition, laptop reviews do not include sentences holding different aspect classes for the same category's identical aspect for distinct aspect words. Next, the laptop dataset has several sentences with multiple aspect categories. Accuracy across the laptop's category is lower when viewed alongside the restaurant reviews. When using a laptop, the aspect extraction task yields result with a lesser degree of precision than recall (there is no phrase with multiple aspect categories for one aspect and no phrase with related aspect categories for different aspects). According to this, in the aspect category recognition test, restaurant reviews have higher precision as opposed to laptop feedback, although the recall for the laptop domain is higher. When ATE and ACD subtasks are completed prior to coreference resolution, accuracy and recall are considerably increased. The laptop dataset contains various intra-sentence relationships as well as a range of aspect subcategories [24]. It is important to note that most reviews in the used datasets only contain one target phrase, and majority objective phrases are stated by a single word. The target can be efficiently extracted using CNNs as a result. The analysis also demonstrates that RNNs have equivalent strength when combined with its other variations. Such combinations can compensate for RNN's inability to recognize crucial phrases.

9. Discussion and major challenges of deep learning-based ABSA

In recent years, aspect extraction's effectiveness has improved significantly using deep learning approaches. Still, there are several issues in a variety of applications where improvement in DL methods is expected. For example, in recent years, the prevailing methods have revealed *domain adaptation* issues for performance degradation if the subject matter of the use of the aspect extraction model is different from the data that was used for training the model. Nonetheless, a huge sample corpus for every task domain must train the correct model to capture the appropriate contextual aspects. Moreover, hand labeling of huge labels for the deep learning model is costly and time-consuming for the distinctive domains [135]. However, attempts have been made in studies such as [118,129,137,174] to extract cross-domain aspects. Therefore, the most effective DL models for ATE subtasks train in a specific domain using independently labeled data. In distinct domains, customers use inconsistent words to convey their emotions. Every so often, interchangeable expressions might express similar sentiments for various domains. For example, in the review, "the story of the film is not predictable", the "predictable" holds a positive sentiment; whereas, in the review, "The scooter brake is not predictable", the "predictable" carries a negative sentiment. Besides, the domain distinction between testing and training data is an issue in the aspect extraction subtask, as the most irrelevant and infrequent sentiment word in one domain might be most relevant and common in different domains [62]. Therefore, the ABSA model trained in a specific domain usually performs unsatisfactorily for other datasets.

However, recent model's performance relies heavily on identical representations of words for a particular term across domains using stable word embedding. Though phrases in the review may be semantically dependent on one another, presented techniques recognize aspect categories for every single sentence individually. In a legitimate ATE challenge, it would be difficult to determine the relationship among various nouns and the overall context of several phrases using a multi-sentence review. The effectiveness of the dataset, which has a variety of aspects and categories, is lowered when an erroneous aspect category is obtained. Therefore, to associate components in review sentences, context and domain expertise are crucial. Additionally, it is still quite difficult to link the context with the aspect phrases from a study of many sentences. Furthermore, it is difficult to extract the domain-specific, most pertinent aspect phrases since there is a lack of contextual and domain information. Because some important aspect terms are not extracted by the present techniques, long-term dependencies for noun phrases cannot be captured. Although deep learning approaches can locate pertinent n-grams, the performance of the aspect category is further hampered by the inability of the current methodologies to capture inter-sentence relationships for accurate multi-word aspect extraction [73].

Table 12
Ambiguity in aspect category detection due to inter-sentence relations.

	Review		Aspect category (Predicted)	Aspect category (Actual)
1	*S* 1:	Tasty dinner, sparkling lights, and pleasant service.	[RESTAURANT#GENERAL]	[FOOD#QUALITY] [AMBIENCE#GENERAL] [SERVICE#GENERAL]
	S 2:	We suggest it!		
2	*S* 1:	Pizza is really amazing.		
	S 2:	Can't avoid tasting it.	[RESTAURANT#GENERAL]	[FOOD#QUALITY]
3	*S* 1:	The staff there is very attentive and down to earth.		
	S 2:	I loved it and would go again.	[RESTAURANT#GENERAL]	[SERVICE#GENERAL]

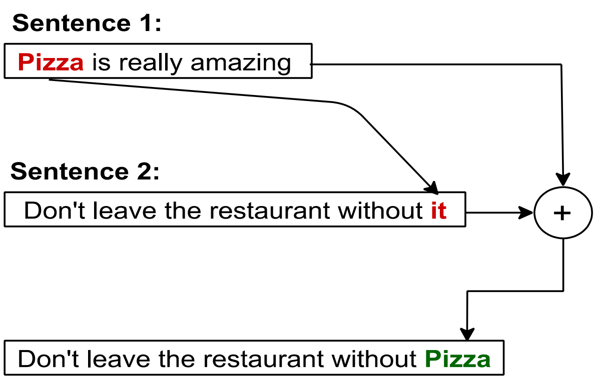

Fig. 19. Example of coreference resolution to improve aspect category detection.

The baseline techniques could not correlate valid multi-word aspects extracted from different review sentences, thus extracting invalid aspect terms. Invalid aspect extraction leads to degrading aspect extraction performance [73]. The current review's sentiment gets evaluated by collecting information on the correlation between sentences and the sentiment of the surrounding sentences. Using the context to match the noun mentioned in the preceding phrase can help in determining the precise significance of the statement. The current machine learning techniques cannot understand the review context [11].

Furthermore, incorporating conceptual and contextual information to associate the context and aspects of a multi-sentence review is still a big challenge [175]. Table 12 gives examples of ambiguity in aspect category detection due to inter-sentence relationships. Contextual knowledge cannot be understood by the current learning strategies, where review contains co-references between sentences. A profound conceptual awareness of the co-reference among phrases is necessary for the relevant analysis of a comprehensive review to determine significant aspect terms, detect aspect categories, and determine sentiment polarity. For a long review, establishing a link between distinct terms and the context of diverse phrases would be difficult to extract a valid multi-word aspect term [18]. The BERT representation and supervised hierarchical attention-based model could be helpful in aspect extraction. Before performing ABSA on multi-sentence reviews, a sentence co-reference resolution step could be added as preprocessing. Therefore, the dataset, which contains a range of aspect phrases and aspect classes, performs worse when an erroneous aspect category is extracted.

As shown in Fig. 19, coreference resolution is one way of retrieving all the nouns and noun phrases that refer to a similar entity in the textual review. The co-reference resolution approaches broadly categorize themselves into linguistic-based and machine learning-based methods (Hobbs' algorithm, centering for pronoun resolution, bridging references, etc.) [176]. In this survey, we have studied and analyzed the DNN models for reviews only in English. Lo et al. [177] provide evidence that multilingual aspect extraction has immense challenges such as word sense ambiguity, the language-dependent structure of the sentence, and translation issues. The model's efficiency differs considerably across languages; hence, one approach that performs well in one language may not guarantee to produce the same performance in other languages. Therefore, lots of effort is required to decode and encode the formal language types. DNN models require very few language-dependent features; still, the performance of ABSA is not yet at its best for multilingual application domains [178]. One of the potential reasons for lower performance in a multilingual environment is due to the unavailability of a sufficient number of review datasets except for SemEval-16 in other languages like Hindi [179], Chinese [180], etc.

The survey observed that the best-performing deep learning models require manually labeled data and domain-task-specific resources in a particular language. The studies show that one promising issue in achieving significant results is small training datasets. According to [105], computation time during the training process is another major issue for most recent DL-based ABSA approaches. The effectiveness of aspect-level sentiment categorization has to be improved by transferring information from the context of coarse-grained document-level work to the domain of fine-grained aspect-level sentiment classification tasks using transfer learning techniques. The deep learning methods for ABSA need to handle the user-generated unstructured text, which does not follow language constraints. The existing methods establish syntactic relations that adhere to grammatical regulations and linguistic restrictions. Still, when posting comments about a product online, users tend to disregard grammatical conventions. The misspellings, abbreviations, sarcasm, humor, common sense, and concept knowledge of the sentences are still unexplored in most existing deep learning methods.

10. Conclusion

For enterprises as well as research societies, the adaption of user-generated text from social media platforms and online portals has offered new communication dimensions. Many attempts to uncover hidden information for analyzing and comprehending sentence meaning have been made over the past 20 years. Deep learning has made a significant contribution to the shifting dynamics in the market by retrieving consumer sentiment through recommendations in a fine-grained manner. These survey' analysis of major deep neural networks has provided a comprehensive overview of aspect-level sentiment analysis approaches. This study provided precise analysis and categorization of more than 80 techniques based on the algorithm and models used for aspect extraction and aspect category detection. The most common ATE and ACD approaches are Convolutional Neural networks, Recurrent Neural networks, long short-term memory, and Gated recurrent units. Some other variants are bidirectional RNNs, hybrid unsupervised rule-based methods, attention mechanisms, cross-domain, hierarchical approaches, etc.

Recently, the models have applied contextual BERT embedding to extend the performance of existing systems. This study of deep learning models has shown some new promising directions in ABSA which are still unexplored. These are domain adaptation, multilingual issues, co-referencing resolution to establish intra-sentence and inter-sentence dependencies, correlating aspects in multi-sentence reviews, linguistic complications (grammatical errors, sarcasm), transfer learning, etc. The existing approaches separately performed ATE and ACD subtasks; still, jointly extracting both aspect extraction and aspect category detection has not achieved optimal results. Hence, it is still a big challenge for research communities. A potential guideline for the near future may be to include common sense, a knowledge base, and concept-specific data into existing deep learning-based ABSA techniques.

Declaration of competing interest

With the submission, authors confirm no conflict of interest and also confirm that the manuscript has not been published nor under consideration else.

Data availability

No data was used for the research described in the article.

Funding

The authors declare that they have not taken any financial support for the conduct of the research and/or preparation of the article.

References

[1] M. Hu, B. Liu, Mining and summarizing customer reviews, in: KDD-2004 - Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2004, pp. 168–177, http://dx.doi.org/10.1145/1014052.1014073.

[2] B. Liu, Sentiment Analysis and Opinion Mining, Morgan & Claypool Publishers, 2012.

[3] Ganpat Singh Chauhan, P. Agrawal, Y.K. Meena, Aspect-based sentiment analysis of students' feedback to improve teaching–learning process, Smart Innov. Syst. Technol. 107 (2019) 259–266, http://dx.doi.org/10.1007/978-981-13-1747-7_25.

[4] Ganpat Singh Chauhan, Y.K. Meena, YouTube video ranking by aspect-based sentiment analysis on user feedback, Adv. Intell. Syst. Comput. 900 (2019) 63–71, http://dx.doi.org/10.1007/978-981-13-3600-3_6.

[5] H. Jangid, S. Singhal, R.R. Shah, R. Zimmermann, Aspect-based financial sentiment analysis using deep learning, in: Companion Proceedings of the the Web Conference 2018, 2018, pp. 1961–1966, http://dx.doi.org/10.1145/3184558.3191827.

[6] K. Du, F. Xing, E. Cambria, Incorporating multiple knowledge sources for targeted aspect-based financial sentiment analysis, ACM Trans. Manage. Inf. Syst (2023) http://dx.doi.org/10.1145/3580480.

[7] S. E, L. Yang, M. Zhang, Y. Xiang, Aspect-based financial sentiment analysis with deep neural networks, in: Companion Proceedings of the the Web Conference 2018, 2018, pp. 1951–1954, http://dx.doi.org/10.1145/3184558.3191825.

[8] W. Medhat, A. Hassan, H. Korashy, Sentiment analysis algorithms and applications: A survey, Ain Shams Eng. J. 5 (4) (2014) 1093–1113, http://dx.doi.org/10.1016/j.asej.2014.04.011.

[9] A.K.J. Ain, Classification of text documents, 41, 8, 1998.

[10] J.R. Quinlan, Induction of decision trees, 2007, pp. 81–106.

[11] K. Schouten, F. Frasincar, Survey on aspect-level sentiment analysis, IEEE Trans. Knowl. Data Eng. 28 (3) (2016) 813–830, http://dx.doi.org/10.1109/TKDE.2015.2485209.

[12] B. Pang, L. Lee, Opinion mining and sentiment analysis, Found. Trends Inf. Retriev. 2 (2) (2008).

[13] K. Ravi, V. Ravi, A survey on opinion mining and sentiment analysis: Tasks, approaches and applications, Knowledge-Based Syst. 89 (2015) 14–46, http://dx.doi.org/10.1016/j.knosys.2015.06.015.

[14] D. Wu, H. Wang, ReviewMiner: An aspect-based review analytics system, in: Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval, 2017, pp. 1285–1288, http://dx.doi.org/10.1145/3077136.3084148.

[15] Zhiyuan Chen, B. Liu, Mining topics in documents: Standing on the shoulders of big data, in: Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2014, pp. 1116–1125, http://dx.doi.org/10.1145/2623330.2623622.

[16] M. Pontiki, D. Galanis, H. Papageorgiou, I. Androutsopoulos, S. Manandhar, M. Al-Smadi, M. Al-Ayyoub, Y. Zhao, B. Qin, O. De Clercq, V. Hoste, M. Apidianaki, X. Tannier, N. Loukachevitch, E. Kotelnikov, N. Bel, S. María Jiménez-Zafra, G. Eryiğit, SemEval-2016 Task 5: Aspect Based Sentiment Analysis. Retrieved May 12, 2020, from http://alt.qcri.org/semeval2014/task4/.

[17] T.A. Rana, Y.N. Cheah, Aspect extraction in sentiment analysis: Comparative analysis and survey, Artif. Intell. Rev. 46 (4) (2016) 459–483, http://dx.doi.org/10.1007/s10462-016-9472-z.

[18] G. Singh Chauhan, Y. Kumar Meena, D. Gopalani, R. Nahta, A two-step hybrid unsupervised model with attention mechanism for aspect extraction, Expert Syst. Appl. 161 (2020) 113673, http://dx.doi.org/10.1016/j.eswa.2020.113673.

[19] Ganpat Singh Chauhan, Y.K. Meena, D. Gopalani, R. Nahta, A mixed unsupervised method for aspect extraction using BERT, Multimedia Tools Appl. 81 (22) (2022) 31881–31906, http://dx.doi.org/10.1007/s11042-022-13023-7.

[20] M. Pontiki, H. Papageorgiou, D. Galanis, I. Androutsopoulos, J. Pavlopoulos, S. Manandhar, Semeval-2014 task 4: Aspect based sentiment analysis, 2014, http://alt.qcri.

[21] L. Dong, F. Wei, C. Tan, D. Tang, M. Zhou, K. Xu, Adaptive recursive neural network for target-dependent Twitter sentiment classification, in: 52nd Annual Meeting of the Association for Computational Linguistics, ACL 2014 - Proceedings of the Conference, Vol. 2, no. 2018, 2014, pp. 49–54, http://dx.doi.org/10.3115/v1/p14-2009.

[22] J.S. Kessler, J.S. Kessler, M. Eckert, L. Clark, N. Nicolov, The ICWSM 2010 JDPA sentiment corpus for the automotive, in: D. Klein, C. Manning (Eds.), Domain International AAAI Conference on Weblogs and Social Media Data Challenge Workshop, 2010, Accurate Unlexicalized Parsing. Retrieved July 12, 2020, from http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.385.3871.

[23] M. Maia, S. Handschuh, A. Freitas, B. Davis, R. McDermott, M. Zarrouk, A. Balahur, WWW'18 Open Challenge, Companion of the Web Conference 2018 on the Web Conference 2018 - WWW '18, 2018, pp. 1941–1942, http://dx.doi.org/10.1145/3184558.3192301.

[24] G. Singh Chauhan, Y. Kumar Meena, D. Gopalani, R. Nahta, A two-step hybrid unsupervised model with attention mechanism for aspect extraction, Expert Syst. Appl. (2020) 113673, http://dx.doi.org/10.1016/j.eswa.2020.113673.

[25] T.A. Rana, Y.N. Cheah, A two-fold rule-based model for aspect extraction, Expert Syst. Appl. 89 (2017) 273–285, http://dx.doi.org/10.1016/j.eswa.2017.07.047.

[26] J. Li, Y. Zhao, Z. Jin, G. Li, T. Shen, Z. Tao, C. Tao, SK2: Integrating implicit sentiment knowledge and explicit syntax knowledge for aspect-based sentiment analysis, in: Proceedings of the 31st ACM International Conference on Information & Knowledge Management, 2022, pp. 1114–1123, http://dx.doi.org/10.1145/3511808.3557452.

[27] M. Polignano, P. Basile, M. Degemmis, G. Semeraro, An emotion-driven approach for aspect-based opinion mining, in: Italian Information Retrieval Workshop, 2018.

[28] B. Liu, W. Hsu, Y. Ma, Integrating classification and association rule mining, 1998, https://www.aaai.org.

[29] S. Blair-Goldensohn, K. Hannan, R. McDonald, T. Neylon, G.a. Reis, J. Reynar, Building a sentiment summarizer for local service reviews, in: WWW Workshop on NLP in the Information Explosion Era, 2008, pp. 339–348, http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Building+a+Sentiment+Summarizer+for+Local+Service+Reviews#0.

[30] A.M. Popescu, O. Etzioni, Extracting product features and opinions from reviews, in: HLT/EMNLP 2005 - Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing, Proceedings of the Conference, 2005, pp. 339–346, http://dx.doi.org/10.3115/1220575.1220618.

[31] S. Li, L. Zhou, Y. Li, Improving aspect extraction by augmenting a frequency-based method with web-based similarity measures, Inf. Process. Manage. 51 (1) (2015) 58–67, http://dx.doi.org/10.1016/j.ipm.2014.08.005.

[32] Y. Kang, L. Zhou, RubE: Rule-based methods for extracting product features from online consumer reviews, Inf. Manag. 54 (2) (2017) 166–176, http://dx.doi.org/10.1016/j.im.2016.05.007.

[33] S. Poria, E. Cambria, L.-W. Ku, C.G. Senticnet, A. Gelbukh, A rule-based approach to aspect extraction from product reviews, 2014, http://alt.qcri.org/semeval2014/task4/index.php?id=data-and-tools.

[34] H. Zhang, Z. Chen, B. Chen, B. Hu, M. Li, C. Yang, B. Jiang, Complete quadruple extraction using a two-stage neural model for aspect-based sentiment analysis, Neurocomputing 492 (2022) 452–463, http://dx.doi.org/10.1016/j.neucom.2022.04.027.

[35] S. Chakraborty, P. Goyal, A. Mukherjee, Aspect-based sentiment analysis of scientific reviews, in: Proceedings of the ACM/IEEE Joint Conference on Digital Libraries in, 2020, 2020, pp. 207–216, http://dx.doi.org/10.1145/3383583.3398541.

[36] H.H. Do, P.W.C. Prasad, A. Maag, A. Alsadoon, Deep learning for aspect-based sentiment analysis: A comparative review, Expert Syst. Appl. 118 (2019) 272–299, http://dx.doi.org/10.1016/j.eswa.2018.10.003.

[37] C. Scaffidi, K. Bierhoff, E. Chang, M. Felker, H. Ng, C. Jin, Red opal: Product-feature scoring from reviews, in: EC'07 - Proceedings of the Eighth Annual Conference on Electronic Commerce, 2007, pp. 182–191, http://dx.doi.org/10.1145/1250910.1250938.

[38] B. Liu, M. Hu, J. Cheng, Opinion observer, in: Proceedings of the 14th International Conference on World Wide Web, WWW '05, 2005, p. 342, http://dx.doi.org/10.1145/1060745.1060797.

[39] L. Zhuang, F. Jing, X.Y. Zhu, Movie review mining and summarization, in: International Conference on Information and Knowledge Management, Proceedings, 2006, pp. 43–50, http://dx.doi.org/10.1145/1183614.1183625.

[40] N. Kobayashi, R. Iida, K. Inui, Y. Matsumoto, Opinion mining on the web by extracting subject-aspect-evaluation relations, in: AAAI Spring Symposium - Technical Report, SS-06-03, 2006, pp. 86–91.

[41] Z. Li, M. Zhang, S. Ma, B. Zhou, Y. Sun, Automatic extraction for product feature words from comments on the web, in: Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 5839 LNCS, 2009, pp. 112–123, http://dx.doi.org/10.1007/978-3-642-04769-5_10.

[42] G. Qiu, B. Liu, J. Bu, C. Chen, Expanding domain sentiment lexicon through double propagation, in: IJCAI International Joint Conference on Artificial Intelligence, 2009, pp. 1199–1204.

[43] S. Raju, P. Pingali, V. Varma, An unsupervised approach to product attribute extraction, in: Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 5478 LNCS, 2009, pp. 796–800, http://dx.doi.org/10.1007/978-3-642-00958-7_88.

[44] W. Jin, H.H. Ho, R.K. Srihari, OpinionMiner: A novel machine learning system for web opinion mining and extraction, in: Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2009, pp. 1195–1203, http://dx.doi.org/10.1145/1557019.1557148.

[45] F. Li, C. Han, M. Huang, X. Zhu, Y.-J. Xia, S. Zhang, H. Yu, Structure-aware review mining and summarization, 2010, http://dx.doi.org/10.5555/1873781.1873855.

[46] N. Jakob, Extracting Opinion Targets in a Single-and Cross-Domain Setting with Conditional Random Fields, no. 11, Association for Computational Linguistics, 2010, http://dx.doi.org/10.5555/1870658.

[47] L. Zhang, S.H. Lim, B. Liu, E. O'brien-Strain, Extracting and ranking product features in opinion documents, 2010, http://dx.doi.org/10.5555/1944566.

[48] W. Xin Zhao, J. Jiang, H. Yan, X. Li, Jointly Modeling Aspects and Opinions with a MaxEnt-LDA Hybrid (Issue 11), Association for Computational Linguistics, 2010, http://dx.doi.org/10.5555/1870658.1870664.

[49] Z. Hai, K. Chang, J.J. Kim, Implicit feature identification via co-occurrence association rule mining, in: Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 6608 LNCS(PART 1), 2011, pp. 393–404, http://dx.doi.org/10.1007/978-3-642-19400-9_31.

[50] J. Yu, Z.-J. Zha, M. Wang, T.-S. Chua, Aspect ranking: Identifying important product aspects from online consumer reviews, 2011, http://thesaurus.com.

[51] K. Bafna, D. Toshniwal, Feature based summarization of customers' reviews of online products, Procedia Comput. Sci. 22 (2013) 142–151, http://dx.doi.org/10.1016/j.procs.2013.09.090.

[52] B. Liu, B. Liu, Sentiment analysis and subjectivity, in: Handbook of Natural Language Processing, second ed., Taylor and Francis Group, Boca, 2010, http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.216.5533.

[53] C. Sauper, R. Barzilay, Automatic aggregation by joint modeling of aspects and values, J. Artificial Intelligence Res. 46 (2013) 89–127, http://dx.doi.org/10.1613/jair.3647.

[54] R. Panchendrarajan, N. Ahamed, B. Murugaiah, P. Sivakumar, S. Ranathunga, A. Pemasiri, Implicit aspect detection in restaurant reviews using cooccurence of words, 2016, pp. 128–136, http://dx.doi.org/10.18653/v1/w16-0421.

[55] B. Ozyurt, M.A. Akcayol, A new topic modeling based approach for aspect extraction in aspect based sentiment analysis: SS-LDA, Expert Syst. Appl. 168 (2020) (2021) 114231, http://dx.doi.org/10.1016/j.eswa.2020.114231.

[56] S. Vashishtha, S. Susan, Highlighting keyphrases using senti-scoring and fuzzy entropy for unsupervised sentiment analysis, Expert Syst. Appl. 169 (2020) (2021) 114323, http://dx.doi.org/10.1016/j.eswa.2020.114323.

[57] A.K. Samha, Y. Li, J. Zhang, Aspect-based opinion mining from product reviews using conditional random fields, 2015, Undefined.

[58] Z. Toh, J. Su, NLANGP at SemEval-2016 Task 5: Improving aspect based sentiment analysis using neural network features. Retrieved May 12, 2020, from https://github.com/JohnLangford/vowpal.

[59] B. Yang, C. Cardie, Extracting Opinion Expressions with semi-Markov Conditional Random Fields, Association for Computational Linguistics, 2012, http://nlp.stanford.

[60] S. Poria, E. Cambria, A. Gelbukh, F. Bisio, A. Hussain, Sentiment data flow analysis by means of dynamic linguistic patterns, IEEE Comput. Intell. Mag. 10 (4) (2015) 26–36, http://dx.doi.org/10.1109/MCI.2015.2471215.

[61] Chuhan Wu, F. Wu, S. Wu, Z. Yuan, Y. Huang, A hybrid unsupervised method for aspect term and opinion target extraction, Knowl.-Based Syst. 148 (2018) 66–73, http://dx.doi.org/10.1016/j.knosys.2018.01.019.

[62] G.S. Chauhan, Y.K. Meena, D. Gopalani, R. Nahta, An unsupervised multiple word-embedding method with attention model for cross domain aspect term extraction, in: Proceedings of 3rd International Conference on Emerging Technologies in Computer Engineering: Machine Learning and Internet of Things, ICETCE 2020, 2020, http://dx.doi.org/10.1109/ICETCE48199.2020.9091738.

[63] B. Wang, H. Wang, Bootstrapping both product features and opinion words from chi-nese customer reviews with cross-inducing, 1.

[64] D.M. Blei, A.Y. Ng, J.B. Edu, Latent Dirichlet allocation Michael I. Jordan, J. Mach. Learn. Res. 3 (2003).

[65] T. Hofmann, Unsupervised learning by probabilistic latent semantic analysis, Mach. Learn. 42 (1–2) (2001) 177–196, http://dx.doi.org/10.1023/A:1007617005950.

[66] L. De Mattei, G. De Martino, A. Iovine, A. Miaschi, M. Polignano, R. Giulia, ATE_ABSITA @ EVALITA2020: Overview of the aspect term extraction and aspect-based sentiment analysis task, 2020, pp. 67–74, http://dx.doi.org/10.4000/books.aaccademia.6849.

[67] S.U.S. Chebolu, P. Rosso, S. Kar, T. Solorio, Survey on aspect category detection, ACM Comput. Surv. 55 (7) (2022) http://dx.doi.org/10.1145/3544557.

[68] Y. Lecun, Y. Bengio, G. Hinton, Deep learning, Nature 521 (7553) (2015) 436–444, http://dx.doi.org/10.1038/nature14539, Nature Publishing Group.

[69] L.M. Rojas-Barahona, Deep learning for sentiment analysis, Lang. Linguistics Compass 10 (12) (2016) 701–719, http://dx.doi.org/10.1111/lnc3.12228.

[70] J. Schmidhuber, Deep learning in neural networks: An overview, Neural Netw. 61 (2015) 85–117, http://dx.doi.org/10.1016/j.neunet.2014.09.003, Elsevier Ltd.

[71] Pengfei Liu, S. Joty, H. Meng, Fine-grained Opinion Mining with Recurrent Neural Networks and Word Embeddings, Association for Computational Linguistics, 2015, https://github.com/ppfliu/opinion-target.

[72] T. Chen, R. Xu, Y. He, X. Wang, Improving sentiment analysis via sentence type classification using BiLSTM-CRF and CNN, Expert Syst. Appl. 72 (2017) 221–230, http://dx.doi.org/10.1016/j.eswa.2016.10.065.

[73] S. Poria, E. Cambria, A. Gelbukh, Aspect extraction for opinion mining with a deep convolutional neural network, Knowl.-Based Syst. 108 (2016) 42–49, http://dx.doi.org/10.1016/j.knosys.2016.06.009.

[74] T. Mikolov, K. Chen, G. Corrado, J. Dean, Efficient estimation of word representations in vector space, in: 1st International Conference on Learning Representations, ICLR 2013 - Workshop Track Proceedings, 2013.

[75] J. Devlin, M.W. Chang, K. Lee, K. Toutanova, BERT: Pre-training of deep bidirectional transformers for language understanding, in: NAACL HLT 2019-2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies - Proceedings of the Conference, 1(Mlm), 2019, pp. 4171–4186.

[76] Y. Goldberg, A primer on neural network models for natural language processing, J. Artificial Intelligence Res. 57 (2016) 345–420, http://www.jair.org/papers/paper4992.html.

[77] D. Tang, F. Wei, B. Qin, N. Yang, T. Liu, M. Zhou, Sentiment embeddings with applications to sentiment analysis, IEEE Trans. Knowl. Data Eng. 28 (2) (2016) 496–509, http://dx.doi.org/10.1109/TKDE.2015.2489653.

[78] Y. Bengio, H. Schwenk, J.-S. Senécal, F. Morin, J.-L. Gauvain, Neural probabilistic language models, in: Innovations in Machine Learning, Springer-Verlag, 2006, pp. 137–186, http://dx.doi.org/10.1007/3-540-33486-6_6.

[79] J. McAuley, J. Leskovec, Hidden factors and hidden topics: Understanding rating dimensions with review text, in: RecSys 2013 - Proceedings of the 7th ACM Conference on Recommender Systems, 2013, pp. 165–172, http://dx.doi.org/10.1145/2507157.2507163.

[80] J. Pennington, R. Socher, C.D. Manning, GloVe: Global vectors for word representation, 2014, http://dx.doi.org/10.3115/V1/D14-1162.

[81] E. Grave, P. Bojanowski, P. Gupta, A. Joulin, T. Mikolov, Learning word vectors for 157 languages, in: LREC 2018-11th International Conference on Language Resources and Evaluation, 2019, pp. 3483–3487.

[82] R. Collobert, J. Weston, J. Com, M. Karlen, K. Kavukcuoglu, P. Kuksa, Natural language processing (almost) from scratch, J. Mach. Learn. Res. 12 (2011) http://dx.doi.org/10.5555/1953048.2078186.
[83] K. Tutar, M.O. Ünalır, L. Toker, Development of a framework for ontology based sentiment analysis on social media, Pamukkale Univ. J. Eng. Sci. 21 (5) (2015) 194–202, http://dx.doi.org/10.5505/pajes.2015.67689.
[84] W. Che, Y. Zhao, H. Guo, Z. Su, T. Liu, Sentence compression for aspect-based sentiment analysis, IEEE/ACM Trans. Audio Speech Lang. Process. 23 (12) (2015) 2111–2124, http://dx.doi.org/10.1109/TASLP.2015.2443982.
[85] S. Jebbara, P. Cimiano, Aspect-based sentiment analysis using a two-step neural network architecture, Commun. Comput. Inf. Sci. 641 (2016) 153–167, http://dx.doi.org/10.1007/978-3-319-46565-4_12.
[86] S. Ruder, P. Ghaffari, J.G. Breslin, INSIGHT-1 at SemEval-2016 task 5: Deep learning for multilingual aspect-based sentiment analysis, in: SemEval 2016-10th International Workshop on Semantic Evaluation, Proceedings, 2016, pp. 330–336, http://dx.doi.org/10.18653/v1/s16-1053.
[87] B. Liu, Sentiment analysis: Mining opinions, sentiments, and emotions, 2015, http://dx.doi.org/10.1017/CBO9781139084789, Sentiment Analysis: Mining Opinions, Sentiments, and Emotions, May –367.
[88] W. Wang, S.J. Pan, D. Dahlmeier, X. Xiao, Recursive neural conditional random fields for aspect-based sentiment analysis, in: EMNLP 2016 - Conference on Empirical Methods in Natural Language Processing, Proceedings, 2016, pp. 616–626, http://dx.doi.org/10.18653/v1/d16-1059.
[89] M.E. Peters, M. Neumann, M. Iyyer, M. Gardner, C. Clark, K. Lee, L. Zettlemoyer, Deep contextualized word representations, in: NAACL HLT 2018-2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies - Proceedings of the Conference, Vol. 1, 2018, pp. 2227–2237, http://dx.doi.org/10.18653/v1/n18-1202.
[90] A. Radford, J. Wu, R. Child, D. Luan, D. Amodei, I. Sutskever, Language models are unsupervised multitask learners, 2018.
[91] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A.N. Gomez, Ł. Kaiser, I. Polosukhin, Attention is all you need, in: Advances in Neural Information Processing Systems, 2017-Decem(Nips), 2017, pp. 5999–6009.
[92] A. Zhao, Y. Yu, Knowledge-based systems knowledge-enabled BERT for aspect-based sentiment analysis, 227, 2021, http://dx.doi.org/10.1016/j.knosys.2021.107220.
[93] X. Li, L. Bing, W. Zhang, W. Lam, Exploiting BERT for end-to-end aspect-based sentiment analysis, 2019, pp. 34–41, http://dx.doi.org/10.18653/v1/d19-5505.
[94] M.Z. Liu, F.Y. Zhou, K. Chen, Y. Zhao, Co-attention networks based on aspect and context for aspect-level sentiment analysis, Knowl.-Based Syst. 217 (2021) 106810, http://dx.doi.org/10.1016/j.knosys.2021.106810.
[95] C. Sun, L. Huang, X. Qiu, Utilizing BERT for aspect-based sentiment analysis via constructing auxiliary sentence, in: NAACL HLT 2019-2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies - Proceedings of the Conference, 1, 2019, pp. 380–385.
[96] J. Su, J. Tang, H. Jiang, Z. Lu, Y. Ge, L. Song, D. Xiong, L. Sun, J. Luo, Enhanced aspect-based sentiment analysis models with progressive self-supervised attention learning, Artificial Intelligence 296 (2021) 103477, http://dx.doi.org/10.1016/j.artint.2021.103477.
[97] Xiaodi Wang, M. Tang, T. Yang, Z. Wang, Knowledge-based systems a novel network with multiple attention mechanisms for aspect-level sentiment analysis, Knowl.-Based Syst. 227 (2021) 107196, http://dx.doi.org/10.1016/j.knosys.2021.107196.
[98] T. Young, D. Hazarika, S. Poria, E. Cambria, Recent trends in deep learning based natural language processing [review article], IEEE Comput. Intell. Mag. 13 (3) (2018) 55–75, http://dx.doi.org/10.1109/MCI.2018.2840738.
[99] Y. Ma, H. Peng, E. Cambria, Targeted aspect-based sentiment analysis via embedding commonsense knowledge into an attentive LSTM, 2014, pp. 5876–5883.
[100] Y. Goldberg, Neural network methods for natural language processing, in: Synthesis Lectures on Human Language Technologies, Vol. 10, no. 1, 2017, pp. 1–311, http://dx.doi.org/10.2200/S00762ED1V01Y201703HLT037.
[101] L. Mai, B. Le, Aspect-based sentiment analysis of Vietnamese texts with deep learning, in: Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 10751 LNAI, 2018, pp. 149–158, http://dx.doi.org/10.1007/978-3-319-75417-8_14.
[102] Lamei Xu, J. Lin, L. Wang, C. Yin, J. Wang, Deep convolutional neural network based approach for aspect-based sentiment analysis, 143(ast), 2017, pp. 199–204, http://dx.doi.org/10.14257/astl.2017.143.41.
[103] C. Sutton, A. Mccallum, C. Sutton, A. Mccallum, An introduction to conditional random fields, Found. Trends R Mach. Learn. 4 (4) (2011) 267–373, http://dx.doi.org/10.1561/2200000013.
[104] Y. Kim, Convolutional neural networks for sentence classification, 2014, http://dx.doi.org/10.3115/V1/D14-1181.
[105] X. Gu, Y. Gu, H. Wu, Cascaded convolutional neural networks for aspect-based opinion summary, Neural Process. Lett. 46 (2) (2017) 581–594, http://dx.doi.org/10.1007/s11063-017-9605-7.
[106] P. Ray, A. Chakrabarti, A mixed approach of deep learning method and rule-based method to improve aspect level sentiment analysis, Appl. Comput. Inform. (2019) http://dx.doi.org/10.1016/j.aci.2019.02.002.
[107] J.L. Elman, Distributed representations, simple recurrent networks, and grammatical structure, Mach. Learn. 7 (2–3) (1991) 195–225, http://dx.doi.org/10.1007/bf00114844.
[108] E. Shi, Q. Li, D. Gu, Z. Zhao, A method of weather radar echo extrapolation based on convolutional neural networks a method of weather radar echo extrapolation based on convolutional neural networks, 2018, 2019, http://dx.doi.org/10.1007/978-3-319-73603-7.
[109] K. Cho, B.Van. Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, Y. Bengio, Learning phrase representations using RNN encoder–decoder for statistical machine translation, in: EMNLP 2014-2014 Conference on Empirical Methods in Natural Language Processing, Proceedings of the Conference, 2014, pp. 1724–1734, http://dx.doi.org/10.3115/v1/d14-1179.
[110] Y. Fan, Y. Qian, F. Xie, F.K. Soong, TTS synthesis with bidirectional LSTM based recurrent neural networks, in: Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH, September, 2014, pp. 1964–1968.
[111] I. Goodfellow, Y. Bengio, A. Courville, Deep Learning.
[112] D. Ma, S. Li, X. Zhang, H. Wang, Interactive attention networks for aspect-level sentiment classification, 2017, http://alt.qcri.org/semeval2014/task4/.
[113] D. Bahdanau, K.H. Cho, Y. Bengio, Neural machine translation by jointly learning to align and translate, in: 3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings, 2015.
[114] Xin Wang, Y. Liu, C. Sun, M. Liu, X. Wang, Extended dependency-based word embeddings for aspect extraction, in: Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 9950 LNCS, 2016, pp. 104–111, http://dx.doi.org/10.1007/978-3-319-46681-1_13.
[115] Chao Wu, Q. Xiong, Z. Yang, M. Gao, Q. Li, Y. Yu, K. Wang, Q. Zhu, Residual attention and other aspects module for aspect-based sentiment analysis, Neurocomputing 435 (2021) 42–52, http://dx.doi.org/10.1016/j.neucom.2021.01.019.
[116] C. Yang, H. Zhang, B. Jiang, K. Li, Aspect-based sentiment analysis with alternating coattention networks, Inf. Process. Manage. 56 (3) (2019) 463–478, http://dx.doi.org/10.1016/j.ipm.2018.12.004.
[117] Peng Liu, L. Zhang, J.A. Gulla, Multilingual review-aware deep recommender system via aspect-based sentiment analysis, ACM Trans. Inf. Syst. 39 (2) (2021) http://dx.doi.org/10.1145/3432049.
[118] Z. Yuan, S. Wu, F. Wu, J. Liu, Y. Huang, Domain attention model for multi-domain sentiment classification, Knowl.-Based Syst. 155 (2018) 1–10, http://dx.doi.org/10.1016/j.knosys.2018.05.004.
[119] C. Sindhu, G. Vadivu, Microprocessors and microsystems fine grained sentiment polarity classification using augmented knowledge sequence-attention mechanism, Microprocess. Microsyst. 81 (2020) (2021) 103365, http://dx.doi.org/10.1016/j.micpro.2020.103365.
[120] Y. Tay, L.Anh. Tuan, S. Cheung Hui, Dyadic memory networks for aspect-based sentiment analysis, 2017, http://dx.doi.org/10.1145/3132847.3132936.
[121] J. Yuan, Y. Zhao, B. Qin, T. Liu, Local contexts are effective for neural aspect extraction, Commun. Comput. Inf. Sci. 774 (2017) 244–255, http://dx.doi.org/10.1007/978-981-10-6805-8_20.
[122] Y. Wang, M. Huang, L. Zhao, X. Zhu, Attention-based LSTM for aspect-level sentiment classification, in: EMNLP 2016 - Conference on Empirical Methods in Natural Language Processing, Proceedings, 2016, pp. 606–615, http://dx.doi.org/10.18653/v1/d16-1058.
[123] X. Tan, Y. Cai, J. Xu, H.F. Leung, W. Chen, Q. Li, Improving aspect-based sentiment analysis via aligning aspect embedding, Neurocomputing 383 (2020) 336–347, http://dx.doi.org/10.1016/j.neucom.2019.12.035.
[124] N. Liu, B. Shen, Knowledge-based systems aspect-based sentiment analysis with gated alternate neural network ☆, Knowl.-Based Syst. 188 (2020) 105010, http://dx.doi.org/10.1016/j.knosys.2019.105010.
[125] M.E. Mowlaei, M.S. Abadeh, H. Keshavarz, Aspect-based sentiment analysis using adaptive aspect-based lexicons, Expert Syst. Appl. 148 (2020) 113234, http://dx.doi.org/10.1016/j.eswa.2020.113234.
[126] T. Kajdanowicz, P. Kazienko, Computer speech & language comprehensive analysis of aspect term extraction methods using various text embeddings i, 69, 2021, http://dx.doi.org/10.1016/j.csl.2021.101217.
[127] Xinyi Wang, F. Li, Z. Zhang, G. Xu, J. Zhang, X. Sun, A unified position-aware convolutional neural network for aspect based sentiment analysis, Neurocomputing (2021) http://dx.doi.org/10.1016/j.neucom.2021.03.092.
[128] Xinyi Wang, G. Xu, Z. Zhang, L. Jin, X. Sun, End-to-end aspect-based sentiment analysis with hierarchical multi-task learning, Neurocomputing (2021) http://dx.doi.org/10.1016/j.neucom.2021.03.100.

[129] Y. Ding, J. Yu, J. Jiang, Recurrent neural networks with auxiliary labels for cross-domain opinion target extraction, 2017, Undefined.
[130] X. Li, W. Lam, Deep multi-task learning for aspect term extraction with memory interaction, in: EMNLP 2017 - Conference on Empirical Methods in Natural Language Processing, Proceedings, 2017, pp. 2886–2892.
[131] W. Xue, W. Zhou, T. Li, Q. Wang, MTNA: A neural multi-task model for aspect category classification and aspect term extraction on restaurant reviews.
[132] H. Ye, Z. Yan, Z. Luo, W. Chao, Dependency-tree based convolutional neural networks for aspect term extraction, in: Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 10235 LNAI, 2017, pp. 350–362, http://dx.doi.org/10.1007/978-3-319-57529-2_28.
[133] D. Dai, T. Chen, S. Xia, G. Wang, Z. Chen, Double embedding and bidirectional sentiment dependence detector for aspect sentiment triplet extraction, Knowl.-Based Syst. 253 (2022) 109506, http://dx.doi.org/10.1016/j.knosys.2022.109506.
[134] R.K. Yadav, L. Jiao, M. Goodwin, O.C. Granmo, Positionless aspect based sentiment analysis using attention mechanism [Formula presented], Knowl.-Based Syst. 226 (2021) 107136, http://dx.doi.org/10.1016/j.knosys.2021.107136.
[135] Ganpat Singh Chauhan, Y. Kumar Meena, D. Gopalani, R. Nahta, An unsupervised multiple word-embedding method with attention model for cross domain aspect term extraction, in: 2020 3rd International Conference on Emerging Technologies in Computer Engineering: Machine Learning and Internet of Things, ICETCE, 2020, pp. 110–116, http://dx.doi.org/10.1109/ICETCE48199.2020.9091738.
[136] D. Meškele, F. Frasincar, ALDONAr: A hybrid solution for sentence-level aspect-based sentiment analysis using a lexicalized domain ontology and a regularized neural attention model, Inf. Process. Manage. 57 (3) (2020) 102211, http://dx.doi.org/10.1016/j.ipm.2020.102211.
[137] M. Dragoni, G. Petrucci, A neural word embeddings approach for multi-domain sentiment analysis, IEEE Trans. Affect. Comput. 8 (4) (2017) 457–470, http://dx.doi.org/10.1109/TAFFC.2017.2717879.
[138] H. Fei, T.-S. Chua, C. Li, D. Ji, M. Zhang, Y. Ren, On the robustness of aspect-based sentiment analysis: Rethinking model, data, and training, ACM Trans. Inf. Syst. 41 (2) (2022) http://dx.doi.org/10.1145/3564281.
[139] B. Liang, R. Yin, L. Gui, J. Du, Y. He, R. Xu, Aspect-invariant sentiment features learning: Adversarial multi-task learning for aspect-based sentiment analysis, in: Proceedings of the 29th ACM International Conference on Information & Knowledge Management, 2020, pp. 825–834, http://dx.doi.org/10.1145/3340531.3411868.
[140] J.A. Kumar, T.E. Trueman, E. Cambria, Gender-based multi-aspect sentiment detection using multilabel learning, Inform. Sci. 606 (2022) 453–468, http://dx.doi.org/10.1016/j.ins.2022.05.057.
[141] H. Zhuang, F. Guo, C. Zhang, L. Liu, J. Han, Joint aspect-sentiment analysis with minimal user guidance, in: Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval, 2020, pp. 1241–1250, http://dx.doi.org/10.1145/3397271.3401179.
[142] C. Musto, P. Lops, M.de. Gemmis, G. Semeraro, Justifying recommendations through aspect-based sentiment analysis of users reviews, in: Proceedings of the 27th ACM Conference on User Modeling, Adaptation and Personalization, 2019, pp. 4–12, http://dx.doi.org/10.1145/3320435.3320457.
[143] P. Pastore, A. Iovine, F. Narducci, G. Semeraro, A general aspect-term-extraction model for multi-criteria recommendations (long paper), 2021, KaRS/ComplexRec@RecSys.
[144] S. Akhtar, T. Garg, A. Ekbal, Neurocomputing multi-task learning for aspect term extraction and aspect sentiment classification, Neurocomputing 398 (2020) 247–256, http://dx.doi.org/10.1016/j.neucom.2020.02.093.
[145] Q. Xu, L. Zhu, T. Dai, C. Yan, Neurocomputing aspect-based sentiment classification with multi-attention network, Neurocomputing 388 (2020) 135–143, http://dx.doi.org/10.1016/j.neucom.2020.01.024.
[146] Zhuang Chen, T. Qian, Retrieve-and-edit domain adaptation for End2End aspect based sentiment analysis, IEEE/ACM Trans. Audio Speech Lang. Process. 30 (2022) 659–672, http://dx.doi.org/10.1109/TASLP.2022.3146052.
[147] K. Srividya, A. Mary Sowjanya, NA-DLSTM – A neural attention based model for context aware aspect-based sentiment analysis, Mater. Today: Proc. xxxx (2021) http://dx.doi.org/10.1016/j.matpr.2021.01.782.
[148] X. Li, L. Bing, P. Li, W. Lam, Z. Yang, Aspect term extraction with history attention and selective transformation*, 2017, https://github.com/lixin4ever/HAST.
[149] Y. Lv, F. Wei, L. Cao, S. Peng, J. Niu, S. Yu, C. Wang, Neurocomputing aspect-level sentiment analysis using context and aspect memory network, Neurocomputing 428 (2021) 195–205, http://dx.doi.org/10.1016/j.neucom.2020.11.049.
[150] Z. Zhou, F. Liu, Neurocomputing filter gate network based on multi-head attention for aspect-level sentiment classification, Neurocomputing 441 (2021) 214–225, http://dx.doi.org/10.1016/j.neucom.2021.02.041.
[151] F. Ren, L. Feng, D. Xiao, M. Cai, S. Cheng, DNet: A lightweight and efficient model for aspect based sentiment analysis, Expert Syst. Appl. 151 (2020) 113393, http://dx.doi.org/10.1016/j.eswa.2020.113393.
[152] Q. Liu, H. Zhang, Y. Zeng, Z. Huang, Z. Wu, Content attention model for aspect based sentiment analysis, in: Proceedings of the 2018 World Wide Web Conference, 2018, pp. 1023–1032, http://dx.doi.org/10.1145/3178876.3186001.
[153] L.C. Cheng, Y.L. Chen, Y.Y. Liao, Aspect-based sentiment analysis with component focusing multi-head co-attention networks, Neurocomputing 489 (2022) 9–17, http://dx.doi.org/10.1016/j.neucom.2022.03.027.
[154] W. Ke, J. Gao, H. Shen, X. Cheng, Neurocomputing, Neurocomputing 456 (2021) 394–406, http://dx.doi.org/10.1016/j.neucom.2021.05.078.
[155] R. Liu, J. Cao, N. Sun, L. Jiang, Aspect feature distillation and enhancement network for aspect-based sentiment analysis, in: Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval, 2022, pp. 1577–1587, http://dx.doi.org/10.1145/3477495.3531938.
[156] B. Zhang, X. Li, X. Xu, K.-C. Leung, Z. Chen, Y. Ye, Knowledge guided capsule attention network for aspect-based sentiment analysis, IEEE/ACM Trans. Audio Speech Lang. Process. 28 (2020) 2538–2551, http://dx.doi.org/10.1109/TASLP.2020.3017093.
[157] B. Xing, I.W. Tsang, Out of context: A new clue for context modeling of aspect-based sentiment analysis, J. Artificial Intelligence Res. 74 (2022) 627–659, http://dx.doi.org/10.1613/JAIR.1.13410.
[158] R. Mokhosi, C. Shikali, Z. Qin, Q. Liu, Maximal activation weighted memory for aspect based sentiment analysis, Comput. Speech Lang. 76 (2020) (2022) 101402, http://dx.doi.org/10.1016/j.csl.2022.101402.
[159] F. Chen, Z. Yang, Y. Huang, A multi-task learning framework for end-to-end aspect sentiment triplet extraction, Neurocomputing 479 (2022) 12–21, http://dx.doi.org/10.1016/j.neucom.2022.01.021.
[160] R. Zhang, Q. Chen, Y. Zheng, S. Mensah, Y. Mao, Aspect-level sentiment analysis via a syntax-based neural network, IEEE/ACM Trans. Audio Speech Lang. Process. 30 (2022) 2568–2583, http://dx.doi.org/10.1109/TASLP.2022.3190731.
[161] J. Zhou, J. Xiangji, Q. Vivian, L. He, Knowledge-based systems SK-GCN: Modeling syntax and knowledge via graph convolutional network for aspect-level sentiment classification, Knowledge-Based Syst. 205 (2020) 106292, http://dx.doi.org/10.1016/j.knosys.2020.106292.
[162] M.P. Geetha, D.K. Renuka, International journal of intelligent networks improving the performance of aspect based sentiment analysis using fi ne-tuned bert base uncased model, Int. J. Intell. Netw. 2 (March) (2021) 64–69, http://dx.doi.org/10.1016/j.ijin.2021.06.005.
[163] Z. Gao, A.O. Feng, X. Song, X.I. Wu, Target-dependent sentiment classification with BERT, 2019, p. 7.
[164] M. Hoang, J. Rouces, Aspect-based sentiment analysis using BERT, 2019.
[165] Z. Zhao, M. Tang, W. Tang, C. Wang, X. Chen, Graph convolutional network with multiple weight mechanisms for aspect-based sentiment analysis, Neurocomputing 500 (2022) 124–134, http://dx.doi.org/10.1016/j.neucom.2022.05.045.
[166] Z. Ren, Q. Shen, X. Diao, H. Xu, A sentiment-aware deep learning approach for personality detection from text, Inf. Process. Manage. 58 (3) (2021) 102532, http://dx.doi.org/10.1016/j.ipm.2021.102532.
[167] Lingling Xu, W. Wang, Improving aspect-based sentiment analysis with contrastive learning, Nat. Lang. Process. J. 3 (2023) 100009, http://dx.doi.org/10.1016/j.nlp.2023.100009.
[168] P. Li, P. Li, X. Xiao, Aspect-pair supervised contrastive learning for aspect-based sentiment analysis, Knowl.-Based Syst. 274 (2023) 110648, http://dx.doi.org/10.1016/j.knosys.2023.110648.
[169] G. Zhao, Y. Luo, Q. Chen, X. Qian, Aspect-based sentiment analysis via multitask learning for online reviews, Knowl.-Based Syst. 264 (2023) 110326, http://dx.doi.org/10.1016/J.KNOSYS.2023.110326.
[170] T. Zhou, Y. Shen, K. Chen, Q. Cao, Hierarchical dual graph convolutional network for aspect-based sentiment analysis, Knowl.-Based Syst. (2023) 110740, http://dx.doi.org/10.1016/j.knosys.2023.110740.
[171] H. Wu, Y. Gu, S. Sun, X. Gu, Aspect-based opinion summarization with convolutional neural networks, in: Proceedings of the International Joint Conference on Neural Networks, 2016-*Octob*, 2016, pp. 3157–3163, http://dx.doi.org/10.1109/IJCNN.2016.7727602.
[172] A. Chaudhuri, S.K. Ghosh, Sentiment analysis of customer reviews using robust hierarchical bidirectional recurrent neural network, Adv. Intell. Syst. Comput. 464 (2016) 249–261, http://dx.doi.org/10.1007/978-3-319-33625-1_23.
[173] K. Schouten, O. van der Weijde, F. Frasincar, R. Dekker, Supervised and unsupervised aspect category detection for sentiment analysis with co-occurrence data, IEEE Trans. Cybern. (2017) http://dx.doi.org/10.1109/TCYB.2017.2688801.

[174] A. García-Pablos, M. Cuadros, G. Rigau, W2VLDA: Almost unsupervised system for aspect based sentiment analysis, Expert Syst. Appl. 91 (2018) 127–137, http://dx.doi.org/10.1016/j.eswa.2017.08.049.
[175] N. Dilawar, H. Majeed, Sentence vector representation methods for aspect category detection. pp. 1–10.
[176] M. Joshi, O. Levy, D.S. Weld, L. Zettlemoyer, BERT for coreference resolution: Baselines and analysis, in: EMNLP-IJCNLP 2019-2019 Conference on Empirical Methods in Natural Language Processing and 9th International Joint Conference on Natural Language Processing, Proceedings of the Conference, 2020, pp. 5803–5808, http://dx.doi.org/10.18653/v1/d19-1588.
[177] S.L. Lo, E. Cambria, R. Chiong, D. Cornforth, Multilingual sentiment analysis: From formal to informal and scarce resource languages, Artif. Intell. Rev. 48 (4) (2017) 499–527, http://dx.doi.org/10.1007/s10462-016-9508-4.
[178] M.S. Akhtar, P. Sawant, S. Sen, A. Ekbal, P. Bhattacharyya, Improving word embedding coverage in less-resourced languages through multilinguality and cross-linguality: A case study with aspect-based sentiment analysis, ACM Trans. Asian Low-Resour. Lang. Inf. Process. 18 (2) (2018) http://dx.doi.org/10.1145/3273931.
[179] M.S. Akhtar, D. Gupta, A. Ekbal, P. Bhattacharyya, Feature selection and ensemble construction: A two-step method for aspect based sentiment analysis, Knowl.-Based Syst. 125 (2017) 116–135, http://dx.doi.org/10.1016/j.knosys.2017.03.020.
[180] J. Feng, S. Cai, X. Ma, Enhanced sentiment labeling and implicit aspect identification by integration of deep convolution neural network and sequential algorithm, Cluster Comput. 22 (3) (2019) 5839–5857, http://dx.doi.org/10.1007/s10586-017-1626-5.