



A mixed unsupervised method for aspect extraction using BERT

Ganpat Singh Chauhan¹ · Yogesh Kumar Meena¹ · Dinesh Gopalani¹ · Ravi Nahta¹

Received: 12 April 2021 / Revised: 24 February 2022 / Accepted: 28 March 2022 /

Published online: 11 April 2022

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2022

Abstract

With the increase of unstructured text on social media platforms from user opinions, deep neural network techniques have significantly contributed to the aspect extraction subtask of Aspect-Based Sentiment Analysis (ABSA). In a multi-sentence review, sentences are contextually interdependent, and static word embedding generates similar representations for the same word in different domains. Hence, existing techniques cannot capture inter-sentence dependencies for valid multi-word aspect extraction. Further, incorporating conceptual information to associate the context and aspect terms is still a challenging task. Therefore, this paper aims to remove inadequate information and capture aspect co-referencing by adding a sentence coreference resolution step before performing ABSA in an unsupervised rule-based method. Next, domain irrelevant aspects are pruned out using contextual embedding. Furthermore, aspects extracted using unsupervised way are given as labeled in training the hierarchical attention-based network using pre-trained language model BERT, Bidirectional Encoder Representations from Transformers. The experimental results on the SemEval-16 dataset show that F-score results are between 2.5% and 5% better than recent supervised deep learning approaches for laptop and restaurant domains, respectively.

Keywords Aspect extraction · BERT · Deep neural network · Sentiment analysis · Unsupervised learning

✉ Ganpat Singh Chauhan
ganpatchauhan@gmail.com; 2018rcp9063@mnit.ac.in

Yogesh Kumar Meena
ymeena.cse@mnit.ac.in

Dinesh Gopalani
dgopalani.cse@mnit.ac.in

Ravi Nahta
2018rcp9077@mnit.ac.in; ravi.nahta@bennett.edu.in

¹ Department of Computer Science & Engineering, Malaviya National Institute of Technology, Jaipur, Rajasthan, India

1 Introduction

In the last decade, online platforms became the primary source of data available in textual forms like reviews, blogs, etc., for consumers, businesses, and researchers [14, 35]. This textual data covers various application domains such as products, education, movies [48], restaurants, government policies, and reviews on online video repository as YouTube [4–6]. These reviews are a valuable source for producers to improve the quality of the product [34]. Still, an overall opinion analysis does not extract the exact sentiment towards different aspects (e.g., food quality, service, or price) [16, 39]. Therefore, in recent times, Aspect-Based Sentiment Analysis (ABSA) approaches have provided excellent results for a variety of application domains [36, 39]. ABSA establishes a correlation between each aspect with the contextual meaning of the sentence [36, 44]. In the reviews of the restaurant domain, ‘Nice Indian food with reasonable price’, the user holds an overall positive opinion (‘nice’ and ‘reasonable’) towards aspect terms: ‘Indian food’ and ‘price’ of the entity ‘FOOD’ as shown in Fig. 1. The words ‘food’ and ‘price’ represent the explicit aspect terms in the above example [13, 36, 44]. In the review: ‘The restaurant is very expensive’, the term ‘price’ is implicitly inferred from the word ‘expensive’ [21, 45].

Rule-based techniques are widely used in sentiment analysis literature for aspect extraction and category detection due to their unsupervised nature. Still, rule-based techniques’ efficiency depends on the performance of the sentences’ language constraints or grammatical correctness [46]. This makes it difficult for the dependency parser to extract multi-word aspects. Moreover, according to [32], the availability of domain-related knowledge is a major concern in capturing the most prominent implicit aspects. [43, 60] explores that although machine learning-based techniques enhanced the aspect extraction efficiency these methods still require lots of manual feature engineering and huge domain-related dataset to train the accurate model. In the last few years, deep neural network-based techniques such as Convolutional Neural Network (CNN), Recurrent Neural Network (RNN), Long Short-Term Memory (LSTM), and its variants, such as [31, 58], etc., have shown significant improvement in aspect extraction. These approaches that easily conceal syntactic and semantic attributes of the words (using word-embedding) can extract intra-sentence relations [10, 25].

Nowadays, a single review might contain several sentences connected via Elaboration (inter-sentence) and Background (intra-sentence) relations according to Rhetorical Structure Theory (RST) [9, 42]. In Elaboration, the context is crucial to extract the sentence’s exact meaning by correlating the noun referred to in the previous sentence. As presented in [4], the



Fig. 1 An example sentence for ABSA

current machine learning techniques are incapable of understanding the review context, where the review contains co-reference between sentences. It would be challenging to capture a correlation between various nouns and the context of multiple sentences for the valid aspect terms extraction in a multi-sentence review. Therefore, obtaining an invalid aspect category degrades the dataset's performance, which has diverse aspects and categories [31]. As in the example multi-sentence review with sentences S1 and S2 as below:

S1: Great food, great décor and great service.

S2: I recommend it!

The predicted aspect category is *[RESTAURANT#GENERAL]*, whereas the actual aspect categories are *[FOOD#QUALITY]*, *[AMBIENCE#GENERAL]*, and *[SERVICE#GENERAL]*. The discussion context gets changes over the sentences in the review; however, there might be an ambiguous context in multi-word aspect extraction [42]. For example, the sentences 'The movie plot is unpredictable' and 'unpredictable steering' holds common opinion words 'unpredictable'. The word 'unpredictable' is positive in the movie domain, but it is obstructive for the aspect *steering*. Aspect extraction is a diversified task with various domains, which require different contextual word embeddings for the same word according to other areas. The recent deep learning-based aspect extraction models perform better when trained on a more massive amount of datasets. However, one of the considerable challenges in ABSA is the shortage of training data for domain-specific tasks.

Deep learning approaches using static word embeddings (word2vec, Glove, etc.) generate the similar embedding for the same word for varieties of domains; hence, they heavily rely on in-domain labeled data [10]. [8] presented that, the shallow representations of words include only previous information in the first layer of deep neural networks. The network's remaining layers are always required to be learned separately for each domain. The existing approaches only extract the semantic meaning of the nouns and noun phrases, thus incapable of retrieving multi-sentence dependencies for aspect terms, which fails to extract valid aspects. Next, the sentence, 'The laptop screen has great resolution' contains the aspect term 'screen resolution' that requires previous and next information to correlate screen with resolution. The context is crucial to extract the sentence's exact meaning by linking the noun referred 'Don't forget to buy *it*' to the previous sentence, 'Pizza is really amazing'. The model uses complete context from all the reviews while correlating long-term dependency words. In text-level aspect extraction, this domain-specific dynamic word embedding helps capture hidden aspects from multiple sentences. Recently, the research community of natural language processing has repeatedly developed techniques using transfer learning to bridge the gap in data representation using a new language representation model called BERT [8, 33].

Our work's motivation is to include a new unsupervised rule-based approach to extract aspects from the sentences of online portals, which doesn't limit users to casually expressing an opinion and violating language rules. A coreference resolution is performed before performing ABSA to remove ambiguity while capturing inter-sentence dependencies in extracting valid aspects. Next, pruning of domain-specific single word and multi-word nouns and noun phrases using dynamic embedding. Moreover, these aspects extracted using unsupervised ways then supplied as labeled for the BERT-based supervised hierarchical attention model. The pre-trained language representation (BERT) with a fine-tuned method increases the ABSA's performance from multiple review sentences.

The hierarchical attention-based network model incorporated the contextual knowledge of the sentences by capturing semantic meanings of the words in varieties of domains; thus, understanding the full context of the review by correlating the inter-sentence aspects of the sentences of a review. Our results validate the impact of contextual word embedding in the aspect term extraction on the SemEval-16 dataset¹ compared to the baselines and most recent approaches. The rest of the paper is organized as follows: Section 2 discussed related work on deep learning strategies for aspect extraction. Next, in Section 3, we have discussed the proposed BERT-based methodology. Moreover, Section 4 analyzed the BERT's effectiveness on aspect extraction, and finally, Section 5 concluded the work.

2 Related work

ABSA has added exceptionally to the analysis of the vast online text in decision-making [4, 47]. The main objectives of ABSA are the aspect term extraction (ATE) and aspect category detection (ACD) subtasks to identify the targets and categories of user opinions from the sentence as given in [39, 44]. The existing ABSA techniques are mainly rule-based methods, machine learning-based techniques, and deep learning-based approaches [10]. The rule-based methods such as series of rules [13], a frequency count of noun/noun phrases [18], calculating *Point-wise Mutual Information* [3, 20], language-model-based method, lexicon-based model [30], etc. can extract explicit aspect from the review. Still, these approaches cannot extract implicit aspects, thus failing to capture contextual knowledge of the review [15]. Further, aspect extraction performance using rule-based methods builds upon the accuracy of the parsers and taggers. Furthermore, these methods rely on the correctness of the language constraints of the sentence.

Therefore, rule-based methods missed extracting some valid noun and noun phrases as aspect terms. Still, aspect category detection performance degrades due to a lack of common-sense and contextual information [10]. Next, [38] applied topic modeling using Latent Dirichlet Allocation (LDA), which outperformed linguistic-based models. [26] further improved the aspect extraction accuracy with the new topic modeling-based method 'sentence segment LDA' for short sentences. These topic modeling approaches can capture document-level aspects but fail to extract aspects at a fine-grained level. These approaches require manual evaluation for unlabeled topics.

Supervised learning methods such as maximum entropy, conditional random field, and support vector machine have improved performance with a sequence labeling problem [43, 60] for aspect extraction. These supervised methods are capable of establishing a correlation between intra-sentence nouns and noun phrases as aspects using contextual information. These supervised models are simple and efficient; they still produce unsatisfactory results when aspect extraction needs the full context of the sentence. Moreover, the manual feature engineering restricts finding the correct context of the review due to inter-sentence dependency, which degrades the system in detecting the right aspect category [10, 47, 58].

[10] has presented that deep neural network models such as RNN, CNN, LSTM, and bi-LSTM have enhanced the aspect extraction results of supervised models by capturing the

¹ <http://alt.qcri.org/semeval2016/task5/index.php?id=data-and-tools>

sentence's semantic meaning. Some of the latest approaches such as a deep CNN combined with new rules [31]; a combined word representation using POS vector and embedded vector as the input to the LSTM based network [21]; a hybrid approach to filter the most relevant domain-specific aspects [58]; an attention-based hybrid model [47]; a sentence vector representation method for word vectors and then used to train the multi-layer deep network [9]; the sentence compression process to convert the complicated sentence into a more natural and shorter sentence before performing aspect extraction and aspect category detection tasks [45]. A variety of LSTM-based deep network models were applied for aspect extraction and category detection on the Laptop and Restaurant domains of SemEval-16 datasets [10, 29]. Some of the recent and top-performing aspect extraction deep learning approaches are as a multi-task learning neural network model [7]; a coupled multi-layer attention model [23]; Dyadic Memory Network model [53]; RNN + WE + Name List + Headword [54]; Hierarchical LSTM based model [17].

In recent times, [56] proposed a unified position-aware CNN model and further improved the model with hierarchical multi-task learning [57]. [24] proposed a hybrid solution for sentence-level ABSA using regularized neural attention model. [40] used BERT-based sentence-level semantic extraction for personality detection considering sentiment information. [52] improved aspect extraction accuracy using pre-trained aspect alignment embedding; still limited to sentence-level aspect extraction on single sentence review. In the last couple of years, several attention-based models such as alternative co-attention model for target-level and context-level for learning effective context representation [61], context-based co-attention network [22], a neural context-aware neural attention network (NA-LSTM) [49], an iterative self-supervised attention based aspect extraction [50], residual attention mechanism [59] are proposed to establish the semantic correlation between the domain context and for each aspect word.

The baseline and recent deep-learning aspect extraction approaches are helpful for single sentence reviews such as the SemEval-14 dataset on *restaurant* and *laptop* domains [10]. The existing approaches can extract the intra-sentence dependencies and correlate the context with the correct aspect about which opinion is given in the sentence. Most of the current methods use static word embedding for the same word in a varieties of contexts [8, 41]. For example, 'river bank' and 'bank account' have two different contexts, but the static word-embedding (using word2vec or Glove) generates the same context-independent embedding for the word 'bank'. These shallow representations of words include only previous information as input to the deep neural networks [8]. The remaining layers of the network are always required to be learned from the start for different tasks. Therefore, text-level aspect extraction (multi-sentence reviews) is yet to explore semantic meaning from words having an inter-sentence dependency on SemEval-16 [29]. It is difficult to capture the long-term dependencies of the long review that requires understating the context of all referred nouns of the previous sentence at the place of the personal pronoun [42]. However, it is still a challenging task for researchers to process long multi-sentence reviews for valid aspect extraction.

The main issue for training a model for ABSA tasks is the shortage of domain-specific datasets. The SemEval-16 dataset needs dynamic embedding across different domains for the same word, which make it possible for aspect extraction using a deep learning model to extract hidden implicit semantics of the review [12, 19, 51, 62]. The general pre-trained language representation trained on the large un-annotated fine-tuned dataset bridges the gap in word representation for a new task model for different domains. In

recent times, two pre-trained approaches, such as the feature-based method called ELMo [28] and the fine-tuned system called Generative pre-trained transformer (OpenAI GPT) [33], have shown performance improvement in ATE task. These approaches alleviate the effort to prepare deep learning models to capture contextual information for valid multi-word aspects. Still, these approaches restrict fully utilizing the power of pre-trained representation, more specifically for the fine-tuned technique. The major issue is that existing language models are unidirectional; therefore, they restrict the architecture for pre-training the model. These techniques are useful when used for the review with a single sentence; still, it can be inefficient when applying text-level aspect extraction tasks where a review holds multi-sentences dependency that needs to incorporate the context of the complete review.

The proposed work has used BERT [8, 19] as a contextual word embedding layer to improve the fine-tuning-based approaches for aspect extraction. A hybrid approach is implemented with an unsupervised linguistic rule-based method with a supervised hierarchical attention-based network using BERT to extend the accuracy of extracting the most relevant domain-relevant aspects from multi-sentence review. First, a set L of new linguistic rules (LR) is prepared to extract multi-word aspects from the reviews which doesn't follow language constraints. A coreference resolution step is added before performing ABSA to remove inadequate information. Further, we propose a pre-trained embedded vector to prune domain-irrelevant prominent aspect terms. Furthermore, these aspects extracted using an unsupervised approach are then supplied as labels in training the supervised deep hierarchical attention-based model using BERT representation. The contextual word embedding for language representation in the proposed model improves the aspect extraction task for varieties of domains by capturing the words' semantic meanings by correlating the inter-sentence aspects of a review's sentences. The experimental results on the standard SemEval-16 dataset validate the BERT's impact in the ATE subtask of ABSA.

3 Proposed BERT based methodology

This article proposes a mixed unsupervised with supervised deep neural network approach using BERT for ATE, as represented in Fig. 2. The proposed approach works as follows: first, dependency parser and POS taggers are used to extract NP chunks from the review sentences. Next, a sentence coreference resolution step is applied for sentence alignment before performing ABSA. We then applied the unsupervised technique with rules given in [10, 31, 37, 46] to extract the noun and noun phrases as aspects and their associated opinions from the NP Chunks.

Next, frequency-based and similarity-based pruning is done using BERT based dynamic embedded vector to select relevant domain-specific aspects and prune out irrelevant aspects. Furthermore, aspects extracted using the above-unsupervised technique are used as labels data for training a BERT-based hierarchical attention-based network for aspect extraction.

3.1 Aspect extraction using unsupervised manner

ABSA's essential subtask is the extraction of nouns and noun phrases as aspect terms from multi-sentence reviews. The proposed approach used a Dependency Parser to extract all NP chunks from the review sentences. Nowadays, noun phrases require intra-sentence and inter-

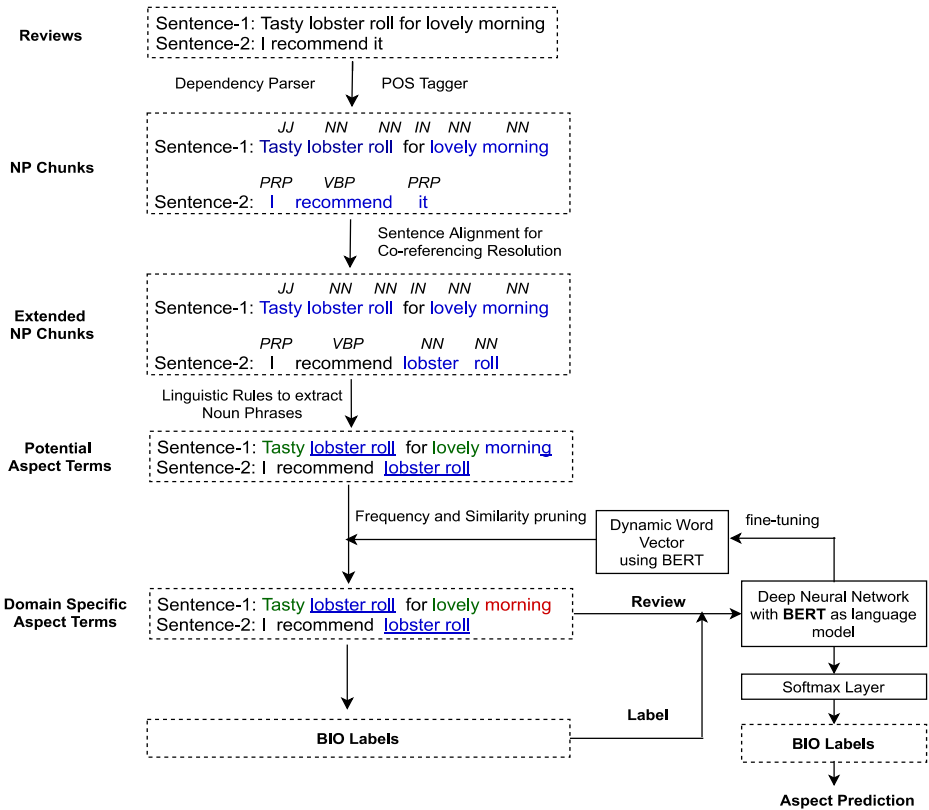


Fig. 2 BERT-based proposed model for ATE

sentence dependencies which depend on the performance of the Parser². In recent times, multi-sentence reviews require co-referencing sentences to avoid ambiguity in extracting the correct aspect terms. Coreference resolution finds all expressions that refer to the same entity in a text, as shown in Fig. 3. This task helps to understand the whole meaning of the sentences for correct aspect extraction in the field of ABSA.

The intra-sentence dependency well captured in recent deep learning techniques of aspect extraction, but the establishment of inter-sentence relation [9, 42] among aspects as per the context of the long sentences captured by replacing the referred noun with the right noun or noun phrase of the previous sentence. In proposed experiments, the ‘*NeuralCoref*’ library of python, integrated with *spaCy*’s used for improving the ambiguity in aspect extraction due to lack of context information. For example, in the review, ‘Pizza is really amazing. I recommend it’, if we use the given review as it is for ABSA, then the predicted aspect category will be *[RESTAURANT#GENERAL]* for sentence-2 of the example review. The inter-sentence relation is essential to understand the review’s context to extract the correct aspect using the detected aspect category. The coreference resolution step converts the review as:

- S1: Pizza is really amazing.
- S2: I recommend Pizza.

² <https://nlp.stanford.edu/software/lex-parser.shtml>

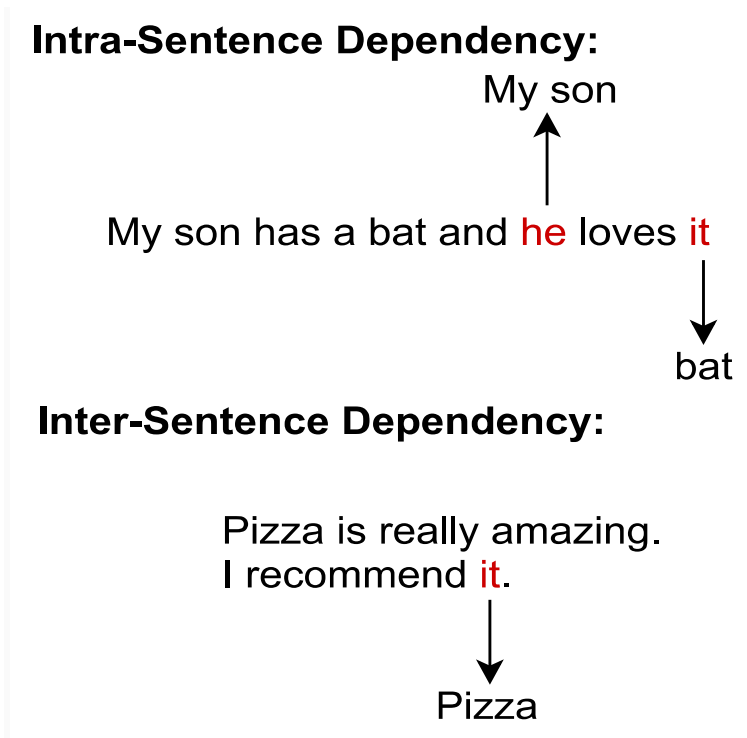


Fig. 3 Examples of sentence-dependencies

After performing the coreference resolution step, each noun/noun phrase of the NP chunks is tagged and associated with the related opinion applying the rules on extracted NP chunks. Nowadays, in multi-sentence reviews, users are unrestricted to follow grammatical rules while writing comments to express their opinions on online portals. Thus, these syntactically incorrect sentences make the Part of speech (POS) tagging task unsafe by missing some valid aspect terms. At the same time, various invalid noun phrases might be extracted as aspect terms from the reviews that don't follow grammatical rules. For example, the sentence: 'great glass of wine', express the opinion 'great' for an aspect 'glass of wine' but the Parser tagged the sentence as:

great/JJ glass/JJ of/IN wine/NN

Here, 'glass' and 'of' are incorrectly marked as adjectives, although they are part of the noun phrase 'glass of wine'. Likewise, in the review: 'Try the crunchy tuna', 'crunchy' tagged as an adjective in place of the noun phrase 'crunchy tuna'.

Try/VB the/DT crunchy/JJ tuna/NN

This paper used noun phrase (NP) chunks to extract single and multi-word aspects and then correlate them with the opinion.

The chunk-level noun phrases of each sentence of the review can reduce the parser issues mentioned above for multi-sentence reviews, which don't follow language rules. According to the dependency tree of one review sentence in Fig. 4 as generated in [46] 'Tasty lobster roll' and 'lovely morning' are two NP chunks. We have created and applied a set L of linguistic rules (LR) from Rule-1 to Rule-8 to extract the most prominent nouns/noun phrases from the extracted NP chunk as given in Algorithm 1. The input to Algorithm 1 for the aspect extraction process: R is a set of tagged reviews, a set of linguistic rules L and a set of opinion words O given by [13] to correlate an adjective with the noun/noun phrase. Here, n represents the noun/noun phrase. The execution process starts with processing reviews (R). The algorithm reads sentences s of each review rev to extract NP chunks c of the sentence. First, if the NP chunk contains any personal pronoun, the coreference resolution (explained above in this section) will be performed.

Furthermore, the linguistic rules from the set L applied to each NP chunk as given in the logic (Rule-1, Rule-2, ..., Rule-8) of Algorithm 1. For each chunk, the algorithm searches any word w_i (noun), and if found, it searches the previous word w_{i-1} as an adjective as per Rule-1. In the example: 'nice ice-cream', nice will be removed. Next, for example, in the NP chunk, 'tasty lobster roll' or in 'try the crunchy tuna'; 'lobster roll' and 'crunchy tuna' will respectively be extracted (Rule-2) to be selected as multi-word prominent aspect terms in set N . In the sentence: 'battery of my laptop', 'my' will be removed and 'battery of laptop' will be extracted as aspect as per Rule-5. Rule-7 overcome the above raise issues in extracting the aspect term 'glass of wine' from the NP chunk: 'great glass of wine'. Accordingly, the other linguistics rules, Rule-2 to Rule-8 given in Algorithm 1, will extract prominent aspects from NP chunks generated from the dependency tree. Rule-8 can easily extract the aspect 'battery backup' from the sentence 'The battery has excellent backup'. The significant enhancement in our aspect extraction process is that we have improved the ambiguity due to personal pronoun using coreference resolution. The review level contextual information has contributed significantly to aspect extraction. The potential aspects using the proposed Algorithm 1 for the example

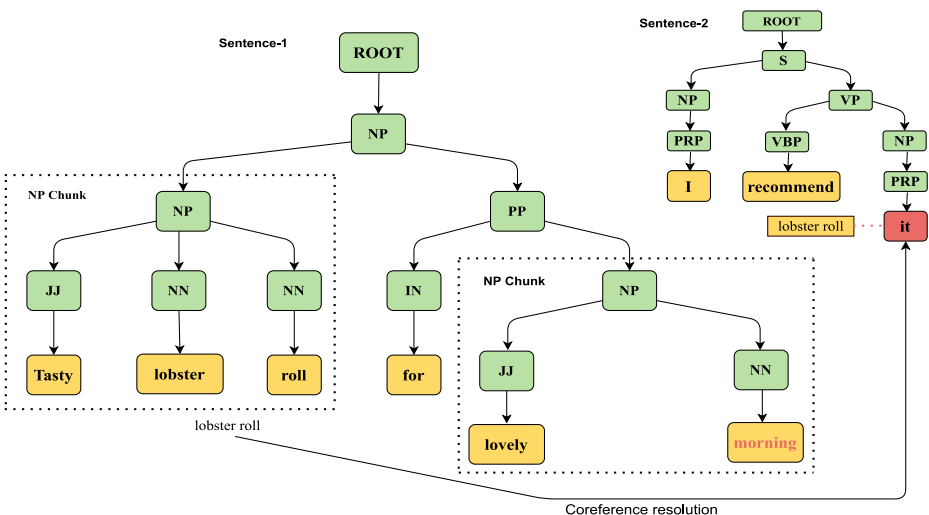


Fig. 4 Coreference resolution on dependency tree of multi-sentence review

given in Figs. 2 and 4 are $\{\textit{lobster roll, morning}\}$. The algorithm outputs a collection N of potential aspects for the review set R .

Algorithm 1 The proposed unsupervised approach for aspect extraction.

Input: a set of reviews (R), a set of rules (L), set of opinion words (O)
Outcome: a set for nouns/noun phrases as aspects (N)

- 1: $N = \{ \}$ // a empty set
- 2: **continue** ((rev)) in (R)
- 3: $D = \{ \}$ // empty Dictionary for sentence-noun pair
- 4: **continue** (s) in rev
- 5: **continue** (c) in s
- 6: **case** c contains personal pronoun
- 7: co-reference resolution using D
- 8: **continue** word (w) in c apply rule from L
- 9: **case** w_i is noun, w_{i-1} is adjective and w_{i-1} is in O // **Rule-1**
- 10: $n = w_i$ // single-word aspect
- 11: **else case** w_i is noun, w_{i-1} is adjective and w_{i-1} is not in O // **Rule-2**
- 12: $n = w_i w_{i-1}$ // multi-word aspect
- 13: **else case** w_i is noun, w_{i-1} is adjective and w_{i-1} is verb // **Rule-3**
- 14: $n = w_i$ // single-word aspect
- 15: **else case** w_i is noun, w_{i-1} is adjective and w_{i-1} is pronoun or article // **Rule-4**
- 16: w_{i-1} need to be removed, if w_{i-1} is the middle, then add left and right of word
- 17: **else case** w_{i-1} is adverb/sentimental verbs/ conjunctions/numerals // **Rule-5**
- 18: w_i will be removed from c
- 19: **else case** w_{i-1} of the noun w_i is sentiment adverbs // **Rule-6**
- 20: remove w_{i-1} from chunk
- 21: **continue** single/multiple word aspects in n ; every n_i associate with any w_{i+1} // **Rule-7**
- 22: add w_{i+1} in n as a noun phrase and update w_i with w_{i+1}
- 23: **end continue**
- 22: **case** w_i belong to O and word (w_{i+1}) as noun // **Rule-8**
- 24: add word w_{i+1} in n and add n in N
- 25: update w_i with w_{i+1}
- 26: **end continue**
- 27: $D[s]=n$
- 28: **end continue**
- 29: **end continue**
- 30: **end continue**
- 31: **Outcome:** a set for nouns/noun phrases as aspects (N)

3.2 Pruning of irrelevant aspects

The set N of extracted noun/noun phrases (from section 3.1) as prominent aspects might contain some irrelevant aspects. These unrelated aspects reduce the accuracy of aspect extraction subtasks. To improve the system's performance by removing irrelevant aspects from set N which are not in task context, we used frequency and similarity pruning process as given in Algorithm 2. Usually, frequently used nouns/noun phrases in the review best represent the particular domain's semantic similarity. Thus, in frequency pruning, the frequency of all the prominent aspects (single word and multi-word) of the set N are computed. The occurrences of the words are calculated for the review, which means if a word occurs in a single review, the frequency will be 1, whereas if the word occurs in 20 reviews, the frequency considered is 20. If a word occurs several times in a multi-sentences of a single review, still, it will be considered as 1. However, the infrequent words and word phrases below the minimum threshold t_f i.e. 2 in our experiments, are pruned out. The performance of varieties of threshold values was tested to decide the threshold level. Next, it is quite possible to have nouns/noun

phrases in the review that don't frequently use, still the most prominent aspects of the task domain.

Algorithm 2 Frequency and similarity pruning process.

Input: Set N extracted in 3.1, set of reviews (R), set of thresholds (T_h), one domain seed word (D)

- 1: $D_A = \{ \}$; Initially D_A contains no aspect
- 2: **continue** word w_i in each aspect $n_i \in N$
- 3: find frequency f_i for every w_i for review R
- 4: $W_n \leftarrow w_i$
- 5: $F_n \leftarrow \text{freq}_{w_i}$
- 6: **end continue**
- 7: **continue** aspect n_i in N
- 8: **case** n_i is a multi-word aspect
- 9: **case** occurrences of word w_i in $n_i \geq$ threshold t_f where $t_f \in T_h$
- 10: $D_A \leftarrow n_i$
- 11: **case** occurrences of word w_i in $n_i <$ threshold t_f and $\text{cos_similarity}(\text{cwe}(n_i), \text{cwe}(D)) > 0.70$
- 12: $D_A \leftarrow n_i$
- 13: **else case** n_i is not a multi-word aspect but $\text{freq}_{n_i} \geq t_f$ where $t_f \in T_h$
- 14: $D_A \leftarrow n_i$
- 15: **else case** n_i is not a multi-word aspect but $\text{freq}_{n_i} < t_f$ and $\text{cos_similarity}(\text{cwe}(n_i), \text{cwe}(D)) > 0.70$
- 16: $D_A \leftarrow n_i$
- 17: **else**
- 18: delete n_i out of N
- 19: delete freq_{n_i} out of f_i within every w_i of n_i
- 20: **end continue**
- 21: **Outcome:** A set D_A for aspect, a set W_n for each words D_A , a set F_n for frequencies of each w_i in W_n

In the next step of pruning, each irregular noun's association is checked by measuring cosine similarity. Therefore, all those prominent aspects pruned in frequency pruning reconsidered if they are contextually similar to the task domain. For this, the aspects whose cosine similarity with the task domain is above the threshold are considered contextually equivalent. In the sentence considered in the example: 'Tasty lobster roll for the lovely morning', the 'Tasty lobster' is selected as a prominent aspect, whereas the infrequent term 'morning' was part of the identified aspect set N is pruned out using frequency pruning. Next, in the similarity pruning process, we reselected 'morning'; still, we have pruned the irrelevant aspect terms 'morning' as its cosine similarity with the task domain was below (0.7 in our experiments) the threshold level. Therefore, the 'morning' aspect will not be considered an aspect term. The pre-trained embedded vector is fined-tuned to update the weights regularly. The BERT-based vector for word representation dynamically generates different embedding vectors [55] for other task domains. Thus, the similarity-based pruning of irrelevant aspects in the proposed approach is much better than word2vec/Glove-based static word embeddings used in [46]. In Eq. 1 for cosine similarity, where $\text{cwe}(n_i)$ represents the contextually embedded weight for word w_i and $\text{cwe}(D)$ is the contextually embedded vector of the one seed word (according to the input reviews) of the task domain is calculated as:

$$\text{cos_similarity}(\text{cwe}(n_i), \text{cwe}(D)) = \frac{\sum_{j=1}^m n_{ij} \cdot D}{\sqrt{\sum_{j=1}^m n_{ij}^2} \sqrt{\sum_{j=1}^m D^2}} \quad (1)$$

In our experiments, we have used the seed word 'Restaurant' in finding cosine similarity when restaurant reviews are taken as input and 'Laptop' seed word for laptop reviews as input. If the task domain is changed to say for Movie reviews in the proposed approach, we need to provide one 'Movie' seed word in calculating cosine similarity.

The methods used in 3.1 and 3.2 don't use labeled data, thus extract the most prominent domain-specific aspects in an unsupervised manner. The pruning process generates D_A as a set of domain-specific prominent aspect terms. The unsupervised approach explained in 3.1, and 3.2 takes the input as a review sentence $R = \{S_1, S_2, \dots, S_N\}$ of N words and generates a set D_A to produce a set of aspect terms $Y = \{Y_1, Y_2, \dots, Y_N\}$ where each Y_i represents as a Begin-Inside-Outside (BIO) label for each word of S_i .

3.3 Training of BERT based proposed deep neural model

A multi-sentence review requires inter-sentence dependency to learn the context using high-level linguistics. Hence, in the proposed ABSA approach, the BERT model is used as the contextual word representation with an attention-based supervised network for better aspect prediction. In this process, the aspect extraction from a given text is framed as a sequential labeling problem. Given an input sentence $R = \{S_1, S_2, \dots, S_N\}$, of N words, the network output will be $Y = \{Y_1, Y_2, \dots, Y_N\}$ and the extracted aspect from an unsupervised method explained in 3.1 and 3.2 are provided as labels (BIO encoding) $Y = \{Y_1, Y_2, \dots, Y_N\}$ to train a BERT-based supervised bi-directional LSTM-based attention model.

3.3.1 Sentence labeling

We have formulated the ATE problem as a sentence labeling task. For sentence labeling, the IOB2 tagging used to encode aspects extracted in an unsupervised way from unlabeled datasets as individual input review labels [46] example given in Table 1. In this tagging, individual word receives the appropriate, feasible tags, namely, **B**, **I** or **O** which indicates **B**egin, **I**nside or **O**utside of the aspect respectively. The word tags in vector representation as:

$$B = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \quad I = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \quad O = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

3.3.2 Features for input representation

In our proposed work, *POS tagging dynamic word embedding with (combined token embedding, position embedding, and segment embedding)* using BERT [55] is used as input features in aspect extraction and aspect category detection. In this paper, contextual word embedding is used from the BERT model as a base for the distributed representation of text in the ABSA task. The BERT uses a multi-layer bidirectional transformer encoder instead of Bi-LSTM to encode syntactical and semantic properties of the textual review. This representation considers the contextual understanding of the word simultaneously from the left and right sides. The transformer block contains a self-attention and a fully-connected layer. BERT's embedding layer that generates the token-level representation using the review's contextual information takes reviews as input.

As represented in Fig. 5a, the input embedding for an individual input sentence is the sum of token embedding, position embedding, and segment embedding. The token embedding is a

Table 1 In BIO labeling the aspects are (lobster roll; morning)

Word	Tasty	Lobster	roll	for	lovely	morning
Y	O	B	I	O	O	B
Type:	Outside	Begin	Inside	Outside	Outside	Begin
Y	[0 0 1]	[1 0 0] [0 1 0]		[0 0 1]	[0 0 1] [0 1 0]	
Aspect:		Aspect-1			Aspect-2	

different embedding for each token, whereas position embedding denotes the token’s position in the review. The segment embeddings for all the sentence tokens will be the same, pointing to the particular sentence to which the token belongs. We have used the generated key features for our aspect extraction approach. Next, in user reviews, POS tags are concatenated as a subsidiary feature for each word. These POS vectors with respective embedding vectors are combined and fed as the input layer to the proposed deep neural network.

3.3.3 Architecture of proposed network model

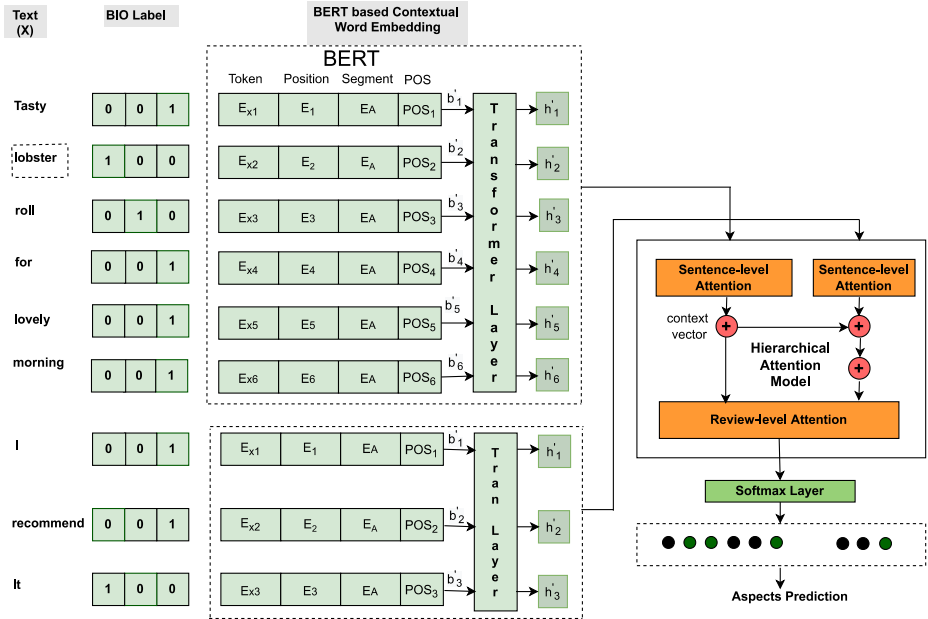
The proposed network architecture given in Fig. 5a shows the one review sentence network of the proposed BERT-based attention model. Suppose a token sequence given for the input sentence is as $R = \{S_1, S_2, \dots, S_N\}$. The formulation of the input embedding is as $R' = \{b'_1, b'_2, b'_3, \dots, b'_N\}$, where b'_i is the combination of token embedding, position embedding, and segment embedding concatenated with its respective POS vector for each input token S_i ; where N denotes number of token in the token sequence. Next, R' passes with L (12 in our experiments) multi-layer transformer blocks to calculate the corresponding contextual representation as $h' = \{h'_1, h'_2, h'_3, \dots, h'_N\} \in \mathbb{R}^N \times dim$ where dim represent the dimension of contextual vector. This POS vector combined contextual representation of the input token sequence (hidden states (768) of the last layer of BERT) used as the embedding layer of the sentence fed to aspect extraction task layers of proposed deep neural networks. Following this BERT-based embedding layer, we used the attention-based model described in [46]. The dropout (0.5) is supplied on the model to reduce overfitting. In our deep learning architecture, given the contextual representation $h' = \{h'_1, h'_2, h'_3, \dots, h'_N\}$ for the review sentence as shown in Fig. 5b, the y_t (a set of BIO tags) will be based on encoder states $H = \{h_1, h_2, \dots, h_N\}$, review-level context (C'_t) and decoder state s_t . The computation of s_t (Eq. 2) is:

$$s_t = f(s_{t-1}, y_{t-1}, C'_t) \tag{2}$$

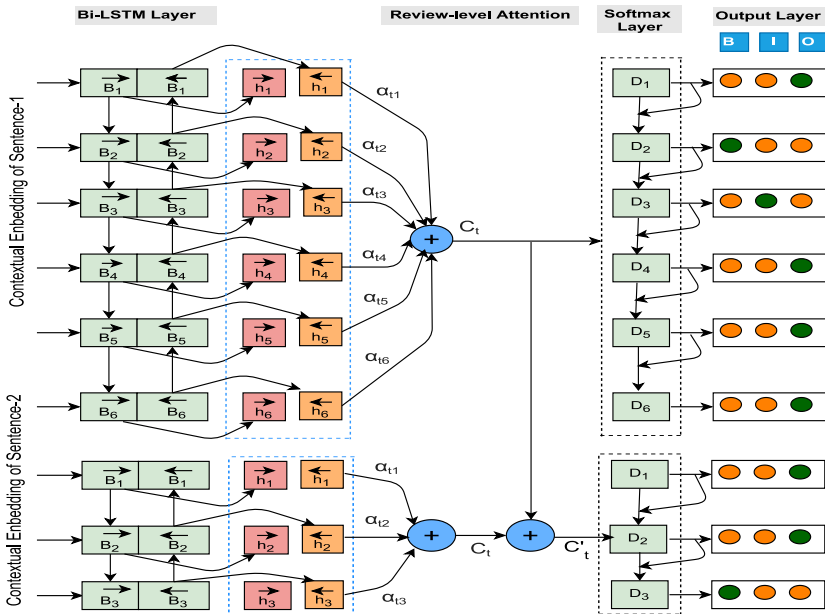
The sentence-level context vector C_t , encoder states h_i and attention weight $\alpha_{ii} = \{\alpha_{t1}, \alpha_{t2}, \dots, \alpha_{tN}\}$ is calculated as:

$$C_t = \sum_{i=1}^N a_{ii} h_i \text{ here } \alpha_{ii} = \{\alpha_{t1}, \alpha_{t2}, \dots, \alpha_{tN}\} \text{ denotes the attention weight} \tag{3}$$

Similar to Eq. 3, context vector for each sentence of the review is calculated. These context-vectors of each sentence is used to evaluate the review-level context C'_t . Here the review-level context C'_t will be calculated using context vector C_t of each sentences as $\sum_{i=1}^n C_t$ where n



(a) Review-level network of proposed model using BERT representation



(b) Hierarchical attention model of proposed review-level network

Fig. 5 a Review-level network of proposed model using BERT representation. b Hierarchical attention model of proposed review-level network

represents number of sentences in a review. Here, the contextual representation h'_i is not directly used in Bi-LSTM network for each hidden state h_i ; instead we first applied

normalization layer [2] when evaluating gates. Next, the probability of the output labels formulated as:

$$P(Y|X) = \pi_{t=1}^N P(y_t | y[1 : t-1], C'_t) \quad (4)$$

where $y[1 : t - 1]$ is equal to $[y_1, y_2, y_3, \dots, y_{t-1}]$, and the probability $P(y_t | y[1 : t-1], C'_t)$ defined as α_{ti} .

A final dense layers D_i in Fig. 5a outputs the probability distribution p_t for each input word of the sentence among **I**, **O**, or **B** tagging scheme. The minimization of categorical cross-entropy loss in the range tagged distribution p_t and expected tags distribution e_t is as:

$$H(e_t, p_t) = - \sum_{k \in K} e_t(k) \log(p_t(k)) \quad (5)$$

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

In proposed techniques using hierarchical review-level attention, the i^{th} label probability of K class classification is a softmax function to compute the attention weight a_{ii} as:

$$\alpha_{ii} = P(y_t = i | X, \theta) = \text{softmax}(e_{ii}) = \frac{\exp(e_{ii})}{\sum_{i=1}^N \exp(e_{ii})} \quad (6)$$

where $e_{ii} = a(s_{t-1}, h_i)$ represent attention energies.

The sentence-level context vector used in finding review-level context for each sentence of the review to capture inter-dependency of aspects in multiple sentences of a review.

4 Evaluation of results

4.1 Dataset analysis

The experimental analysis is measured on the laptop and restaurant reviews in standard SemEval-2016 datasets for the ATE task. All the reviews in training and testing datasets annotated with aspects, aspect category, and polarity. In the review: ‘The pizza is overpriced and soggy’, both the aspects term are ‘pizza’, aspect categories are *FOOD#QUALITY* and *FOOD#PRICES*, and sentiment polarity for both the aspects (pizza) is marked ‘negative’. The restaurant and laptop dataset statistics for training and testing data of SemEval-16 shown in Table 2. The laptop dataset does not have aspects as labeled data. There is a higher variety of aspect categories in the laptop (88) than the restaurant (12) dataset. Both the restaurant and laptop dataset contains multi-sentence reviews. Table 3 describes aspect-based details on varieties of indicators for both sample data of the restaurant domain. The training sample of the dataset holds 35% non-aspect sentences, whereas the testing sample contains 39% of dataset sentences that don’t have any aspect. Moreover, single-word aspects are higher than multi-word aspects in both training and testing data. Further, a single sentence might also contains multiple aspects terms. Hence, the datasets are unbalanced in terms of aspects/non-aspect reviews and reviews with single-word/multi-word aspects.

Table 2 Dataset statistics for proposed approach [46]

Indicator	Restaurant			Laptop		
	Training Set	Testing Set	Gross	Training Set	Testing Set	Gross
Total Samples	395	91	486	451	80	531
Sentences	2000	676	2676	2500	800	3300
Aspect Category	12	12	12	81	68	88

Next, Table 4 shows statistics of SemEval-16 for aspect categories. The maximum sentences in the restaurant reviews discusses about aspects from *'FOOD#QUALITY'* as aspect category, whereas *LAPTOP#GENERAL* is the maximal aspect category mentioned in the laptop reviews. The laptop reviews don't contain any aspect as labeled data. Both the dataset might hold the same category for varieties of aspect terms in a single review. For example, the sentence: 'Sauce was watery, and the food didn't have much flavor', holds the same category *'FOOD#QUALITY'* for the aspect terms 'Sauce', and 'food'. The dataset also has reviews that give a different category for identical aspect terms in a sentence. Furthermore, the datasets also hold cross-sentence dependencies among different nouns and noun-phrases. Therefore, our proposed method of aspect extraction using BERT will be beneficial in performance improvement for reviews of both the considered task domains.

4.2 Evaluation metrics

As presented above in the Table 3, the dataset is unbalanced; thus, in the result evaluation of our proposed model for aspect extraction, precision, recall, and F-score are used as follows:

$$\text{Precision} = (\text{True Positive}) / (\text{True Positive} + \text{False Positive}) \quad (7)$$

$$\text{Recall} = (\text{True Positive}) / (\text{True Positive} + \text{False Negative}) \quad (8)$$

$$f\text{-Score} = (2 * \text{Precision} * \text{Recall}) / (\text{Precision} + \text{Recall}) \quad (9)$$

Table 3 Aspect-based details of Restaurant dataset [46]

Indicator	Training	Testing
Total Reviews	395	91
Total Sentences	2000	676
# of Aspect Sentences	1254	414
# of reviews with Noun as Aspects	898	331
#. of sentence with Nouns Phrase as Aspects	356	83
# of sentences with multiple category	901	347
# of multi-aspect reviews	353	67
Total # of Aspects	2507	651
# of Unique Aspects	635	265
Overall Multi-word aspects	379	135
Overall Noun aspects	256	130

Table 4 Detailed statistics of SemEval-16 for some aspect categories (Restaurant and Laptop)

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			

4.3 Performance comparison

The performance of the proposed hybrid unsupervised technique for aspect extraction combined with BERT-based representation analyzed with benchmark rule-based approaches such as double propagation (DP) [31] and conditional random field (CRF) [43, 60]. The proposed hierarchical attention-based further compared with recent aspect extraction supervised deep learning-based approaches [1, 19, 27, 59] and previous hybrid approaches [46, 58] for aspect extraction. These techniques are helpful when used for the review with a single sentence; still, it can be inefficient when applying text-level aspect extraction tasks where a review holds multi-sentences dependency that needs to incorporate the context of the complete review.

The word2vec/Glove based word embedding (WE) may miss some valid aspects from different parts of the long sentence, containing multiple aspect categories. Table 5 shows the impact of contextual BERT-based language representation as features for the deep network in aspect extraction tasks using coreference resolution. Compared to traditional word embedding concatenate with POS [46], the contextual BERT-based features for individual sentence tokens extended the performance of feature representation. Both precision and recall are improved with and without using the coreference resolution step by inspecting the contextual information for inter-sentence dependency in relevant aspect extraction for restaurant (+6%) and laptop (+4%) reviews. The laptop domain contains a variety of aspect categories compared to the restaurant domain. Here, recall of laptop decreased by 6%, which says several relevant noun phrases missed as aspects. Both the precision and recall increased when the BERT-based model with POS was used as a feature representation. Therefore, as a feature, the BERT

Table 5 Feature analysis of the proposed methodology using BERT

		Without Coreference			With Coreference		
Context	Properties	Precision	Recall	F-Score	Precision	Recall	F-Score
Restaurant	Word Embedding (WE)	81.85	79.84	80.83	82.35	78.36	80.31
	WE + POS	82.54	81.04	81.78	84.54	80.22	82.32
	Contextual BERT	86.22	83.36	84.77	87.12	85.38	86.24
Laptop	Word Embedding (WE)	64.64	77.78	70.60	67.44	78.56	72.58
	WE + POS	76.42	72.24	74.27	78.14	75.86	76.98
	Contextual BERT	78.88	79.76	79.32	80.64	78.44	79.52

correctly marks nouns and noun phrases as valid aspect terms. These features consider inter-sentence dependency to detect the correct aspect category and capture the entire review's contextual information.

The impact of the BERT representation observed at different phases of the proposed model as presented in the flowchart (Fig. 2) is shown in Table 6. The domain-specific aspect extraction performance of the proposed unsupervised approach (linguistic rules (LR) + coreference + pruning) for multi-sentence review has significantly improved up to 14% and 10% for restaurant and laptop domains, respectively. The results show that linguistic rules (LR) applied on NP chunks better extract multi-word aspects from reviews that don't follow language constraints. The coreference resolution step further improves the aspect extraction by removing inadequate information where a review holds multi-sentences dependency that needs to incorporate the context of the complete review. The irrelevant aspect pruning step significantly improves the results and extracts prominent domain-specific aspects. Further, the domain-specific aspects extracted using the unsupervised approach (LR + coreference + pruning) are used as labeled to train the proposed hierarchical attention-based model using BERT representation. Compared with our previous work LR+ Bi-LSTM + Attention [46], the work proposed in this paper improved the performance of aspect extraction. The contextual word embedding for language representation in the proposed model improves the aspect extraction task for varieties of domains by capturing the words' semantic meanings by correlating the inter-sentence aspects of a review's sentences. Our approach minimizes the cost of manual annotation as we have not explicitly applied labels to the model. Although the accuracy of aspect extraction and aspect category increased significantly, as the laptop dataset contains various multi-word aspects, the proposed model gives a lower performance for the laptop dataset than the restaurant dataset.

Here, the label annotation not provided for laptop reviews; still, the proposed rule-based approach with a coreference resolution step benefits to detect the correct aspect category. In our rules-based approach for chunk-based aspects retrieval, the precision is not so promising, which retrieves a variety of unrelated aspects. The lower recall involves a small number of logical aspects, which have intra-sentence and inter-sentence dependency. The laptop reviews hold lesser contextual aspects; thus, the restaurant reviews' precision value is better than the laptop. The coreference model increases precision (9% for laptops and 7% for restaurants). Besides, the recall increases to around 2% for both laptop and restaurant datasets. These results indicate that the laptop dataset contains more reviews with long-term inter-sentence

Table 6 Effect of coreference and BERT in the aspect extraction on SemEval-16 corpus

Context	Model	Precision	Recall	F-Score
Restaurant	LR (potential aspects)	61.66	67.73	64.55
Restaurant	LR+coreference	68.38	70.72	69.53
Restaurant	LR+ coreference+ pruning (Domain-specific Aspects)	79.24	78.66	78.95
Restaurant	Bi-LSTM + Attention + LR [46]	81.02	77.09	79.01
Restaurant	Proposed BERT-based Model	85.16	81.33	83.20
Laptop	LR (potential aspects)	52.96	62.64	57.39
Laptop	LR+coreference	61.64	64.49	63.03
Laptop	LR+ coreference+ pruning (Domain-specific Aspects)	64.49	63.22	63.85
Laptop	Bi-LSTM + Attention + LR [46]	75.34	71.36	73.33
Laptop	Proposed BERT-based Model	68.08	75.98	71.81

dependency between multiple words to extract valid noun phrases. Further, when coreference resolution combined with.

frequency and similarity pruning step to filter most relevant domain-specific aspects, the precision and recall are significantly improved to 11% and 8%, respectively, for the restaurant dataset. Simultaneously, the pruning step for filtering the irrelevant aspect has less impact on precision.

The above results further contribute that the restaurant dataset contains more domain-specific aspects. The degradation of recall in the laptop dataset is due to the domain-dependent aspect terms. The earlier phase’s prominent aspects were supplied as labels to the proposed approach using the BERT-based language model. This BERT-based model improves the prediction of aspect extraction by improving both precision and recall. The proposed model also increases the recall for the laptop dataset. The precision of laptop reviews is low compared to Bi-LSTM + attention model as the laptop reviews hold lesser contextual aspects. Table 6 further reveals that every individual module brings out excessive recall compared to the precision and low precision due to the handy of reviews that contain different aspect categories for the same aspect terms in a sentence. Simultaneously, the reviews carry the same aspect categories for different aspect terms in the restaurant domain. Furthermore, BERT-based network with the coreference resolution step also upgraded the results to detect aspect category for both the domains compared to RNN + WE + Name List + Headword + Word cluster (M6), RNN + Fine-tune + WE (M7), and Bi-GRU + WE + POS (M8) as shown in Fig. 6. In Fig. 6, (P1) represents proposed model using the coreference step without BERT whereas (P2) represents proposed model using coreference step with BERT. The results show that the BERT-based hierarchical attention model can establish the long-term dependency between nouns/noun phrases of different sentences to capture the overall context of the sentences. The F-score tested on various threshold values for BERT-based similarity to remove irrelevant aspects. For all datasets, the similarity score may remain the same, thus average F-score is used to observe the impact. It examined that if the threshold level is below 5, the performance was not sensitive for both restaurant and laptop domains. The performance further improved by extending the attention technique to BERT-based hierarchical attention model to capture all the multi-word aspects which missed due to inter-sentence dependency for erroneous grammatical reviews. Compared to the latest work [46], the precision and recall in the proposed

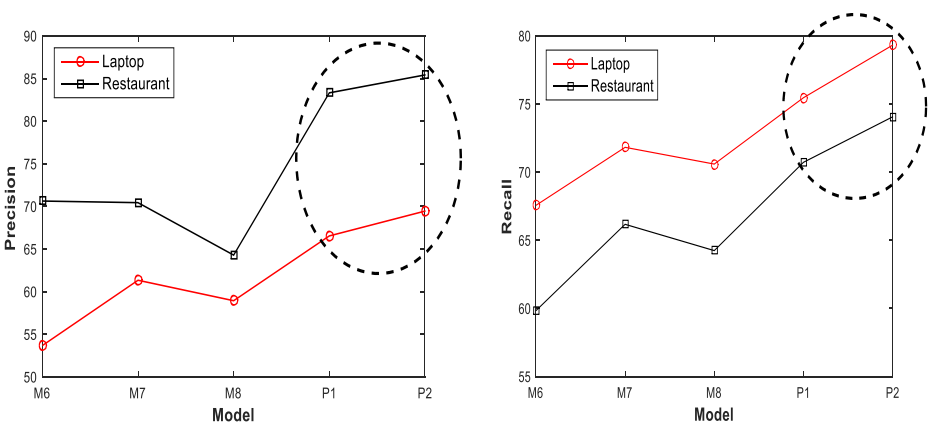


Fig. 6 Effect of precision and recall of BERT-based system for aspect category detection

approach are improved by 4% and 5% for restaurant reviews, whereas the recall for laptop reviews is enhanced by 4%.

As revealed in Fig. 7 and Table 7, our proposed approach using the coreference step without using BERT (P1) accomplished somewhat similar performance with the existing baseline and recent techniques, including [46] for aspect extraction. Besides, our proposed hierarchical attention-based model using coreference step with BERT (P2) improved the performance of precision (68%) and recall (75.98%) for laptop reviews and precision (85.16%) and recall (81.33%) for restaurant reviews as compared to the results achieved in [46] for aspect extraction. The results presented in Table 7 show that the proposed hybrid

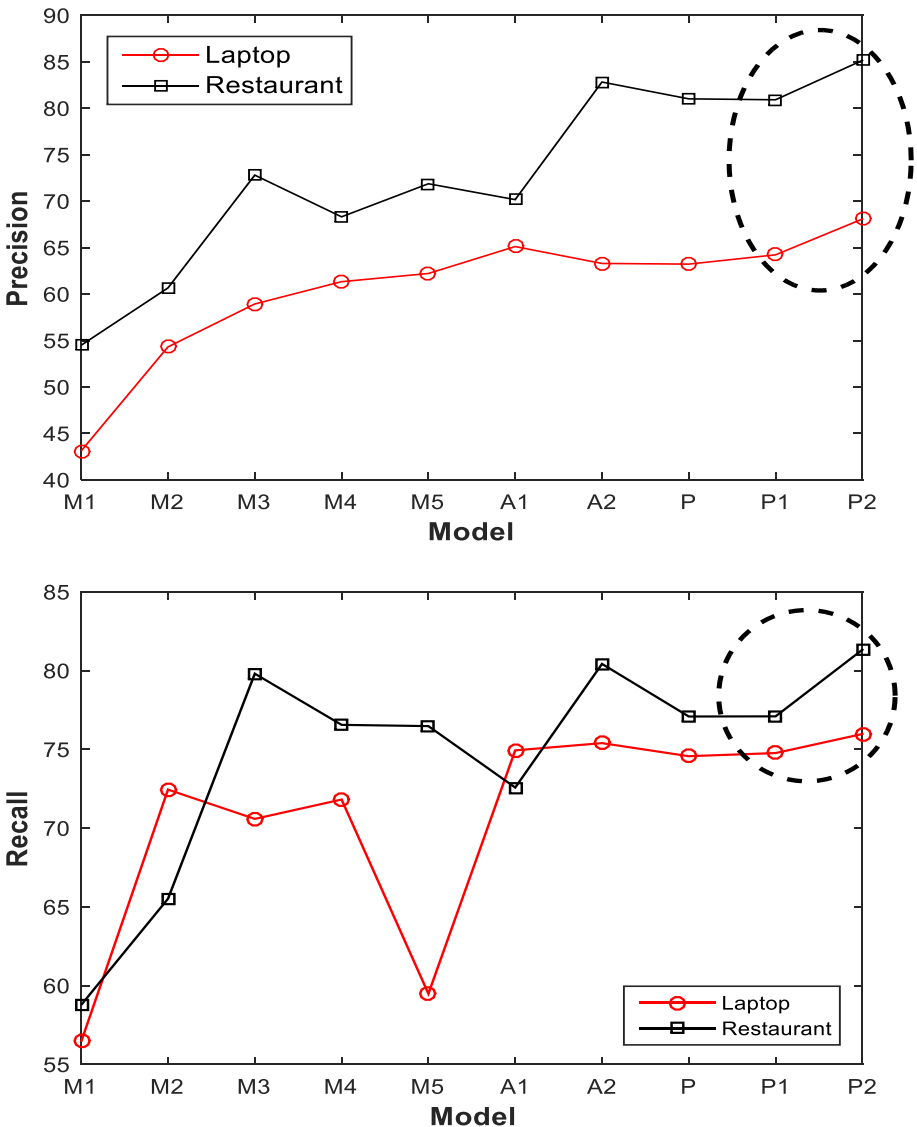


Fig. 7 Effect of precision and recall of proposed BERT-based system for aspect extraction

Table 7 Performance of proposed BERT-based model

Baseline and Proposed Approaches	Laptop			Restaurant		
	Precision	Recall	F-Score	Precision	Recall	F-Score
DP [31] (M1)	43.13	56.45	48.90	54.5	58.78	56.56
CRF [43, 60] (M2)	54.32	72.45	62.09	60.67	65.52	63.00
Bi-GRU+ WE + POS [58] (M3)	58.94	70.59	64.24	72.81	79.81	76.15
Bi-LSTM + Google WE + CRF [7] (M4)	61.34	71.82	66.17	68.32	76.56	72.21
Contextual Embedding + CRF [19] (M5)	62.22	59.49	60.78	71.88	76.48	74.06
Residual Attention (A1) [59]	65.12	74.94	69.72	70.17	72.58	71.34
Bi-LSTM + self-attention (A2) [1]	63.30	75.40	57.26	82.83	80.43	81.61
Unsupervised (LR)+Pruning + Attention [46] (P)	63.23	74.57	68.43	81.02	77.09	79.01
Proposed Hierarchical Attention Model (coreference without BERT) (P1)	64.23	74.77	69.10	80.92	77.10	78.96
Proposed Hierarchical Attention Model (coreference with BERT) (P2)	68.08	75.98	71.81	85.16	81.33	83.20

unsupervised approaches (P1) and (P2) achieved better performance than baseline rule-based methods (M1) and (M2) [31, 43, 60], supervised deep learning-based models (M4), (M5) [7, 19] and attention-based recent models (A1) and (A2) [1, 59], and recent hybrid models (M3), (P) [47, 58] for laptop and restaurant domains. The aspect extraction task produces a lower precision than recall for the laptop (no sentence with varieties of categories for an aspect and no sentence with similar aspect categories for various aspects). Similarly, in the aspect category detection task, the recall for laptop domain is higher, while precision for the restaurant upgraded compared to laptop reviews. Both precision and recall are improved significantly when coreference resolution performed before aspect extraction and aspect category detection tasks. The overall F-score of aspect extraction for the laptop is 11% less than the restaurant reviews (when coreference resolution applied) when BERT applied because it contains a fewer single word and noun phrase aspects, as given in Table 4. The overall F-score of aspect category detection for the laptop is 8% less than the restaurant domain (when coreference resolution applied). The laptop reviews have a variety of aspect categories and several intra-sentence dependencies. Our experiments show that coreference resolution performed before extracting aspects and their categories better captures the syntactic and semantic meaning of the review.

The proposed model don't use labeled data still reach significant results compared to recent approaches for different samples of data. Therefore, our proposed work minimize the dependency of labels data and reduce the cost of manual annotation. The proposed mixed unsupervised approach with supervised BERT weighted attention upgraded the accuracy of domain adaptation issues when the new domain faced without any labeled data in cross-domain aspect extraction. Hence, the proposed model has remarkably handled the context of alteration issues when the new domain faces any labeled data. The result analysis validates the coreference resolution performed before the concatenated POS with the BERT embedding layer better represents the syntactic and semantic understating of the review.

4.4 Test of significance for the proposed model

To prove the consistency of the proposed hybrid model, the F-score is calculated on varieties of sample data. It is elaborated from Fig. 8 that the proposed system's accomplishments on

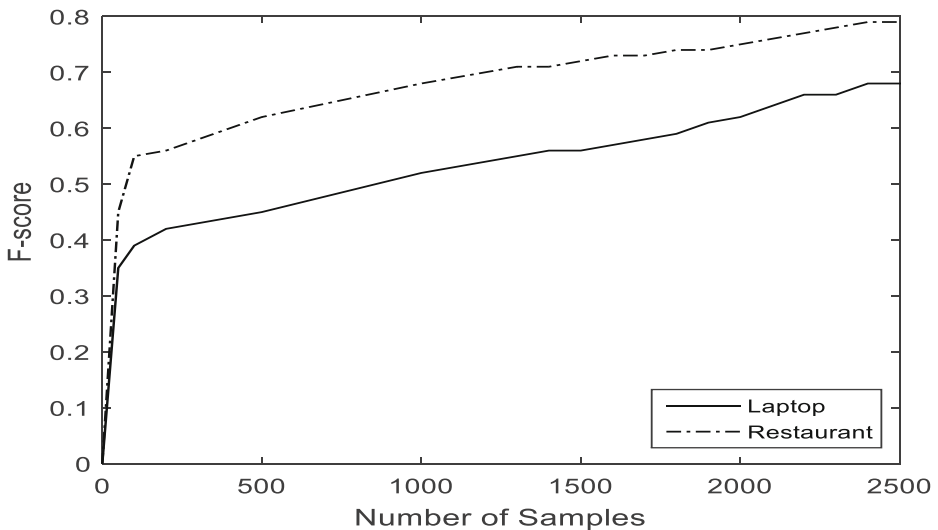


Fig. 8 Evaluation of F-score on varieties of sample data taken from [46]

various data samples are almost equivalent to recent benchmark models. Figure 8 shows that the performance of extended work presented in this paper is equivalent to previous work when a number of samples are about 1700 sentences for restaurant reviews and around 1500 samples for laptop reviews. In addition, the significance test as used in [11] is applied to elaborate that the proposed hybrid model accurately extracts the domain-specific aspects. This test verifies the significance, merits, and consistency of the proposed hybrid model (**P2**) compared to the previous model (**P**) LR+ Bi-LSTM+ Attention [46], thus ensuring the superiority of the proposed model. The actual aspects extracted from proposed unsupervised models (used as labels in the hierarchical attention-based model) are paired with the aspect extracted from the proposed BERT-based hierarchical attention model.

The null hypothesis here is that no significant difference exists between the F-score of the aspect extracted using unsupervised way and the aspect extracted using hierarchical attention-based deep learning model for different data samples. The matching of the F-score of the aspects extracted in the proposed unsupervised manner and aspect extracted using the proposed hierarchical attention-based deep learning model generates the p value (0.4) of the model (**P**) LR+ Bi-LSTM + Attention [46] for different data samples. This p value exceeds the significance level of $\alpha = 0.05$. The p value (0.25) of the proposed hybrid unsupervised model (**P2**) is less than the p value of the model (**P**) [46] for different data samples. Hence, the null hypothesis is correct, which means that there is no significant distinction between the proposed hybrid model and the previous model (**P**) LR + Bi-LSTM + Attention [46]. This statistical significance results elaborate that the proposed hybrid model has smaller errors and aspect extraction accuracy.

4.5 Case study to achieve precision, recall and F-score

This section will discuss a case study to achieve the value of precision, recall, and F-score. This study considered the restaurant reviews from the SemEval-16 dataset, which we used to evaluate the proposed model. In section 4, Tables 2, 3, and 4 are presented to represent the

dataset statistics of the Semeval-16 dataset. The restaurant reviews statistics presented in Table 4 show that the testing samples contain 651 aspects in 91 reviews. The samples shows that the reviews with nouns as aspect are higher than noun phrases as aspects. In addition, the classes are unbalanced in terms of aspects reviews, aspect category, and a number of aspects. Hence, the proposed hybrid model used precision, recall, and F-score as a performance measure in place of accuracy. The focus here is to improve the value of True Positive (TP) and reduce the values of False Negative (FN) and False Positive (FP). The higher TP means the more nouns/noun phrases are extracted as valid aspects. FP represents wrong nouns/noun phrases are extracted as aspects. FN represents that some valid aspects are missed to be extracted as aspects. The confusion matrix produces the values of TP , FP , and FN . These values are then used to calculate the values of precision, recall, and F-score as formulated in Eqs. 7, Eq. 8 and Eq. 9, respectively. Now, consider, step-by-step evaluation of the proposed hybrid unsupervised approach to achieve the precision, recall, and F-score using the values of TP , FP and FN as presented in Table 8. The details and functioning of each step are discussed in Section 3; hence, this section only focuses on performance measures.

Initially, when new linguistic rules (LR) are applied on NP chunks of the review sentence for potential aspects extraction, the values $TP(401)$, $FP(195)$, and $FN(250)$ are measured. These values show that 401 valid aspects are correctly extracted as aspects, 195 nouns/noun phrases which are not potential aspects; still, wrongly extracted as valid aspects. In contrast, 250 valid aspects are missed to be extracted. The higher values of FP and FN represent valid some aspects are missed, and invalid aspects are extracted, respectively. Hence, next, we added the coreference resolution step with (LR) to reduce the values to $FP(179)$ and $FN(219)$ and improve $TP(432)$ value. Further, the pruning step with ($LR +$ coreference) is applied to reduce the values of $FP(140)$ and $FN(135)$ and improve $TP(516)$ value. Furthermore, the BERT-based hierarchical attention model reduces the values of $FP(127)$ and $FN(97)$ and improve $TP(554)$ value. The results show the significance of each step of the proposed hybrid unsupervised model.

5 Conclusion

The adaptation of user-generated text from social networking sites and web portals has opened new dimensions of communication for businesses and research communities. In the last two decades, many attempts have been made to capture hidden information for analyzing and understanding the meaning of the sentence. Extracting the user opinion from reviews in a fine-grained using deep learning has contributed remarkably to the changing dynamics in the marketplace. In the existing aspect extraction approaches, aspect categories for all the sentences are detected independently, although sentences in a review are contextually

Table 8 Values of TP , FP and FN on test samples at different phases of the proposed model

Context	Model	TP	FP	FN
Restaurant	LR (potential aspects)	401	195	250
Restaurant	LR+coreference	432	179	219
Restaurant	LR+ coreference + pruning (Domain-specific Aspects)	516	140	135
Restaurant	Proposed BERT-based Hierarchical Attention Model	554	127	97

interdependent to each other. It is essential to have context and domain knowledge to correlate aspects in different review sentences. Furthermore, associating the context with the aspect terms from a review with multiple sentences is still not explored. However, the existing techniques cannot capture inter-sentence dependencies for valid multi-word aspect extraction, which further leads to degrading the aspect category's performance. We introduce an unsupervised BERT-based attention-based model for ATE tasks in the proposed work. Our approach consists of four subtasks. Initially, a co-referencing resolution step added before aspect extraction, which improves uncertainty due to co-reference between noun phrases and multi-word aspects. Next, a set of linguistic rules are incorporated to detect varieties of multi-word aspects from grammatically incorrect sentences. Further, we propose an embedded weight-based pruning to filter relevant contextual aspect terms. Finally, these aspect terms extracted from an unsupervised manner are provided as label data to train the BERT-based proposed model. The BERT for language model correlates different sentences that extract the contextual meaning of the previous and next sentences of the review and capture aspect dependencies of the multi-sentence review. Our experiment results of the proposed mixed unsupervised approach combined with BERT-based language representation are better than most recent supervised deep learning models in the ATE task on the standard SemEval-16 dataset. Some new promising directions in ABSA that are still unexplored are domain adaptation, multilingual issues, correlating aspects in multi-sentences review, linguistics complications (grammatical errors, sarcasm), etc. Incorporating the common-sense, knowledge base, and concept-specific information in current deep learning-based ABSA approaches may be a challenging task for the near future.

Declarations

Conflict of interest The authors confirm no conflict of interest with the submission and ensure that the manuscript has not been published nor under consideration else.

References

1. Akhtar S, Garg T, Ekbal A (2020) Neurocomputing multi-task learning for aspect term extraction and aspect sentiment classification. *Neurocomputing* 398:247–256. <https://doi.org/10.1016/j.neucom.2020.02.093>
2. Ba JL, Kiros JR, Hinton GE (2016) Layer normalization
3. Blair-Goldensohn S, Hannan K, McDonald R et al (2008) Building a sentiment summarizer for local service reviews. *WWW Work NLP Inf Explos Era*:339–348
4. Chauhan GS, Kumar Meena Y (2018) Prominent aspect term extraction in aspect based sentiment analysis. In: 3rd international conference and workshops on recent advances and innovations in engineering, ICRAIE 2018. Institute of Electrical and Electronics Engineers Inc, pp 1–6. <https://doi.org/10.1109/ICRAIE.2018.8710408>
5. Chauhan GS, Meena YK (2019) YouTube video ranking by aspect-based sentiment analysis on user feedback. In: *Advances in intelligent systems and computing*. Springer Verlag, pp 63–71, vol 900. Springer, Singapore. https://doi.org/10.1007/978-981-13-3600-3_6
6. Chauhan GS, Agrawal P, Meena YK (2019) Aspect-based sentiment analysis of students' feedback to improve teaching–learning process. In: *Smart innovation, Systems and Technologies*. Springer Science and Business Media Deutschland GmbH, pp 259–266, vol 107. Springer, Singapore. https://doi.org/10.1007/978-981-13-1747-7_25
7. Chen T, Xu R, He Y, Wang X (2017) Improving sentiment analysis via sentence type classification using BiLSTM-CRF and CNN. *Expert Syst Appl* 72:221–230. <https://doi.org/10.1016/j.eswa.2016.10.065>
8. Devlin J, Chang MW, Lee K, Toutanova K (2019) BERT: Pre-training of deep bidirectional transformers for language understanding. *NAACL HLT 2019–2019 Conf North Am Chapter Assoc Comput Linguist Hum Lang Technol - Proc Conf* 1:4171–4186
9. Dilawar N, Majeed H Sentence Vector Representation Methods for Aspect Category Detection. 1–10

10. Do HH, Prasad PWC, Maag A, Alsadoon A (2019) Deep learning for aspect-based sentiment analysis: a comparative review. *Expert Syst Appl* 118:272–299. <https://doi.org/10.1016/j.eswa.2018.10.003>
11. Fan GF, Yu M, Dong SQ, Yeh YH, Hong WC (2021) Forecasting short-term electricity load using hybrid support vector regression with grey catastrophe and random forest modeling. *Util Policy* 73:101294. <https://doi.org/10.1016/j.jup.2021.101294>
12. Hoang M, Rouces J (2019) Aspect-based sentiment analysis using BERT
13. Hu M, Liu B (2004) Mining and summarizing customer reviews. In: *KDD-2004 - proceedings of the tenth ACM SIGKDD international conference on knowledge discovery and data mining*. ACM Press, New York, pp 168–177
14. Jebbara S, Cimiano P (2016) Aspect-based sentiment analysis using a two-step neural network architecture. In: *Communications in computer and information science*. Springer Verlag, 641, pp 153–167. <https://doi.org/10.1007/s10994-013-5413-0>
15. Kang Y, Zhou L (2017) RuBE: rule-based methods for extracting product features from online consumer reviews. *Inf Manag* 54:166–176. <https://doi.org/10.1016/j.im.2016.05.007>
16. Kersting JGM (2021) Human language comprehension in aspect phrase extraction with importance weighting. *Nat Lang Process Inf Syst* 12801:231–242. https://doi.org/10.1007/978-3-030-80599-9_21
17. Li X, Lam W (2017) Deep multi-task learning for aspect term extraction with memory interaction, EMNLP. *Association for Computational Linguistics*, pp 2886–2892. <https://doi.org/10.18653/v1/D17-1310>
18. Li S, Zhou L, Li Y (2015) Improving aspect extraction by augmenting a frequency-based method with web-based similarity measures. *Inf Process Manag* 51(1):58–67. <https://doi.org/10.1016/j.ipm.2014.08.005>
19. Li X, Bing L, Zhang W, Lam W (2019) Exploiting BERT for end-to-end aspect-based sentiment analysis. 34–41. <https://doi.org/10.18653/v1/d19-5505>
20. Liu B, Hsu W, Ma Y (1998) Integrating classification and association rule mining
21. Liu P, Joty S, Meng H (2015) Fine-grained opinion mining with recurrent neural networks and word Embeddings. *Association for Computational Linguistics*
22. Liu MZ, Zhou FY, Chen K, Zhao Y (2021) Co-attention networks based on aspect and context for aspect-level sentiment analysis. *Knowledge-Based Syst* 217:106810. <https://doi.org/10.1016/j.knosys.2021.106810>
23. Ma D, Li S, Zhang X, Wang H (2017) Interactive attention networks for aspect-level sentiment classification, *IJCAI, Artificial Intelligence (cs.AI); Computation and Language (cs.CL)*. <https://doi.org/10.48550/arXiv.1709.00893>
24. Meškelė D, Frasincar F (2020) ALDONAR: a hybrid solution for sentence-level aspect-based sentiment analysis using a lexicalized domain ontology and a regularized neural attention model. *Inf Process Manag* 57(3):102211. <https://doi.org/10.1016/j.ipm.2020.102211>
25. Mikolov T, Chen K, Corrado G, Dean J (2013) Efficient estimation of word representations in vector space. In: *1st international conference on learning representations, ICLR 2013 - workshop track proceedings*. *International conference on learning representations, ICLR*
26. Ozyurt B, Akcayol MA (2021) A new topic modeling based approach for aspect extraction in aspect based sentiment analysis: SS-LDA. *Expert Syst Appl* 168:114231. <https://doi.org/10.1016/j.eswa.2020.114231>
27. Peng C, Zhongqian S, Lidong B, Yang W (n.d.) *Recurrent Attention Network on Memory for Aspect Sentiment Analysis*
28. Peters ME, Neumann M, Iyyer M et al (2018) Improving language understanding by. *OpenAI*:1–10
29. Pontiki M, Galanis D, Papageorgiou H, et al (n.d.) *SemEval-2016 Task 5: Aspect Based Sentiment Analysis*
30. Popescu AM, Etzioni O (2005) 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*. *Association for Computational Linguistics, Morristown*, pp 339–346. <https://aclanthology.org/H05-1043>
31. Poria S, Cambria E, Gelbukh A (2016) Aspect extraction for opinion mining with a deep convolutional neural network. *Knowledge-Based Syst* 108:42–49. <https://doi.org/10.1016/j.knosys.2016.06.009>
32. Qiu G, Liu B, Bu J, Chen C (2011) Opinion word expansion and target extraction through double propagation
33. Radford A, Wu J, Child R, et al (2018) *Language models are unsupervised multitask learners*
34. Rajput R, Solanki A (2016) *International journal of computer science and Mobile computing review of sentimental analysis methods using lexicon based approach*. *Int J Comput Sci Mob Comput* 5:159–166
35. Rajput R, Solanki A (2017) Real time sentiment analysis of tweets using machine learning and semantic analysis. *Commun Comput Syst - Proc Int Conf Commun Comput Syst ICCCS 2016*:687–692. <https://doi.org/10.1201/9781315364094-123>
36. Rana TA, Cheah YN (2016) Aspect extraction in sentiment analysis: comparative analysis and survey. *Artif Intell Rev* 46:459–483. <https://doi.org/10.1007/s10462-016-9472-z>
37. Rana TA, Cheah YN (2017) A two-fold rule-based model for aspect extraction. *Expert Syst Appl* 89:273–285. <https://doi.org/10.1016/j.eswa.2017.07.047>

38. Rana TA, Cheah YN, Letchmunan S (2016) Topic modeling in sentiment analysis: a systematic review. *J ICT Res Appl* 10:76–93
39. Ravi K, Ravi V (2015) A survey on opinion mining and sentiment analysis: tasks, approaches and applications. *Knowledge-Based Syst* 89:14–46. <https://doi.org/10.1016/j.knsys.2015.06.015>
40. Ren Z, Shen Q, Diao X, Xu H (2021) A sentiment-aware deep learning approach for personality detection from text. *Inf Process Manag* 58:102532. <https://doi.org/10.1016/j.ipm.2021.102532>
41. Rojas-Barahona LM (2016) Deep learning for sentiment analysis. *Lang Linguist Compass* 10:701–719. <https://doi.org/10.1111/lnc3.12228>
42. Ruder S, Ghaffari P, Breslin JG (n.d.) A Hierarchical Model of Reviews for Aspect-based Sentiment Analysis
43. Samha AK, Li Y, Zhang J (2015) Aspect-based opinion mining from product reviews using conditional random fields. *Undefined*
44. Schouten K, Frasinca F (2016) Survey on aspect-level sentiment analysis. *IEEE Trans Knowl Data Eng* 28: 813–830. <https://doi.org/10.1109/TKDE.2015.2485209>
45. Schouten K, van der Weijde O, Frasinca F, Dekker R (2017) Supervised and unsupervised aspect category detection for sentiment analysis with co-occurrence data. *IEEE Trans Cybern* 48:1263–1275. <https://doi.org/10.1109/TCYB.2017.2688801>
46. Singh Chauhan G, Kumar Meena Y, Gopalani D, Nahta R (2020) A two-step hybrid unsupervised model with attention mechanism for aspect extraction. *Expert Syst Appl* 113673. <https://doi.org/10.1016/j.eswa.2020.113673>
47. Singh Chauhan G, Kumar Meena Y, Gopalani D, Nahta R (2020) A two-step hybrid unsupervised model with attention mechanism for aspect extraction. *Expert Syst Appl* 161:113673. <https://doi.org/10.1016/j.eswa.2020.113673>
48. Singh T, Nayyar A, Solanki A (2020) Multilingual opinion mining movie recommendation system using RNN. Springer, Singapore
49. Srividya K, Mary Sowjanya A (2021) NA-DLSTM – a neural attention based model for context aware aspect-based sentiment analysis. *Mater Today Proc* <https://doi.org/10.1016/j.matpr.2021.01.782>
50. Su J, Tang J, Jiang H, Lu Z, Ge Y, Song L, Xiong D, Sun L, Luo J (2021) Enhanced aspect-based sentiment analysis models with progressive self-supervised attention learning. *Artif Intell* 296:103477. <https://doi.org/10.1016/j.artint.2021.103477>
51. Sun C, Huang L, Qiu X (2019) Utilizing BERT for aspect-based sentiment analysis via constructing auxiliary sentence. *NAACL HLT 2019–2019 Conf North Am Chapter Assoc Comput Linguist Hum Lang Technol - Proc Conf* 1:380–385
52. Tan X, Cai Y, Xu J, Leung HF, Chen W, Li Q (2020) Improving aspect-based sentiment analysis via aligning aspect embedding. *Neurocomputing* 383:336–347. <https://doi.org/10.1016/j.neucom.2019.12.035>
53. Tay Y, Anh Tuan L, Cheung Hui S (2017) Dyadic memory networks for aspect-based sentiment analysis. <https://doi.org/10.1145/3132847.3132936>
54. Toh Z, Su J (2016) NLANGP at SemEval-2016 task 5: improving aspect based sentiment analysis using neural network features
55. Vaswani A, Shazeer N, Parmar N, et al (2017) Attention is all you need. *Adv Neural Inf Process Syst* 2017-Decem:5999–6009
56. Wang X, Li F, Zhang Z, Xu G, Zhang J, Sun X (2021) A unified position-aware convolutional neural network for aspect based sentiment analysis. *Neurocomputing*. 450:91–103. <https://doi.org/10.1016/j.neucom.2021.03.092>
57. Wang X, Xu G, Zhang Z, Jin L, Sun X (2021) End-to-end aspect-based sentiment analysis with hierarchical multi-task learning. *Neurocomputing*. 455:178–188. <https://doi.org/10.1016/j.neucom.2021.03.100>
58. Wu C, Wu F, Wu S, Yuan Z, Huang Y (2018) A hybrid unsupervised method for aspect term and opinion target extraction. *Knowledge-Based Syst* 148:66–73. <https://doi.org/10.1016/j.knsys.2018.01.019>
59. Wu C, Xiong Q, Yang Z, Gao M, Li Q, Yu Y, Wang K, Zhu Q (2021) Residual attention and other aspects module for aspect-based sentiment analysis. *Neurocomputing* 435:42–52. <https://doi.org/10.1016/j.neucom.2021.01.019>
60. Yang B, Cardie C (2012) Extracting opinion expressions with semi-Markov conditional random fields. *Association for Computational Linguistics*
61. Yang C, Zhang H, Jiang B, Li K (2019) Aspect-based sentiment analysis with alternating coattention networks. *Inf Process Manag* 56:463–478. <https://doi.org/10.1016/j.ipm.2018.12.004>
62. Yuan Z, Wu S, Wu F, Liu J, Huang Y (2018) Domain attention model for multi-domain sentiment classification. *Knowledge-Based Syst* 155:1–10. <https://doi.org/10.1016/j.knsys.2018.05.004>