Knowledge-Based Systems 246 (2022) 108668

Contents lists available at ScienceDirect

Knowledge-Based Systems

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

An enhanced guided LDA model augmented with BERT based semantic strength for aspect term extraction in sentiment analysis

Manju Venugopalan, Deepa Gupta *

Department of Computer Science and Engineering, Amrita School of Engineering, Bengaluru, Amrita Vishwa Vidyapeetham, India

A R T I C L E I N F O

Article history:
Received 24 May 2021
Received in revised form 20 March 2022
Accepted 24 March 2022
Available online 31 March 2022

Keywords:
Sentiment analysis
Aspect term extraction
Guided LDA
BERT
Semantic similarity

A B S T R A C T

Aspect level sentiment analysis is a fine-grained task in sentiment analysis. It extracts aspects and their corresponding sentiment polarity from opinionated text. The first subtask of identifying the opinionated aspects is called aspect extraction, which is the focus of the work. Social media platforms are an enormous resource of unlabeled data. However, data annotation for fine-grained tasks is quite expensive and laborious. Hence unsupervised models would be highly appreciated. The proposed model is an unsupervised approach for aspect term extraction, a guided Latent Dirichlet Allocation (LDA) model that uses minimal aspect seed words from each aspect category to guide the model in identifying the hidden topics of interest to the user. The guided LDA model is enhanced by guiding inputs using regular expressions based on linguistic rules. The model is further enhanced by multiple pruning strategies, including a BERT based semantic filter, which incorporates semantics to strengthen situations where co-occurrence statistics might fail to serve as a differentiator. The thresholds for these semantic filters have been estimated using Particle Swarm Optimization strategy. The proposed model is expected to overcome the disadvantage of basic LDA models that fail to differentiate the overlapping topics that represent each aspect category. The work has been evaluated on the restaurant domain of SemEval 2014, 2015 and 2016 datasets and has reported an F-measure of 0.81, 0.74 and 0.75 respectively, which is competitive in comparison to the state of art unsupervised baselines and appreciable even with respect to the supervised baselines.

1. Introduction

Sentiment Analysis [1–3] is a research area inspired by the urge to know what others care or think. Also known as opinion mining, [4,5] it is a text analysis technique that detects sentiment within text. Mass opinion is always a key factor that aids in decision making, be it an individual choosing a travel destination or a company revisiting a product design. Understanding end users' reactions are essential for business since customers express their views more openly than before. Even from a company's perspective, automatically analyzing customer feedback through survey responses and social media conversations creates opportunities for brands to receive inputs from their customers, and tailor products and services accordingly.[1] The applications of sentiment analysis even go real time [6,7] when a company wants to judge brand sentiment on social media, so that unhappy customers are sensed immediately and receive an instantaneous response. Thus the applications of sentiment analysis [8] are boundless. In addition, the huge availability of opinionated data, owing to the growing popularity of social media has accelerated research in the field.

The immense popularity of the field has created a demand for algorithms that automatically extract sentiment from opinionated text. Initial research in this area focused on classifying opinionated data/reviews as positive, negative or neutral [9–11]. Attempts were made to report fine-grained sentiment polarity by expressing the polarity value on a scale of 10 or so, rather than categorical values [12–14]. The initial source of data was mostly product reviews from websites where people express their experience of having used a product. Another active subfield was related to building polarity lexicons [15–17] which elevated the results of sentiment classification. With the advent of social media, the volume of data to work on increased exponentially with more scope for experimentations but came in with equal challenges posed by the informal nature and short length of opinionated texts [18–20]. As the importance of the field grew, the demand was for fine-grained analysis, rather than reporting the sentiment orientation as a whole. A microscopic view of the opinionated text that summarizes review text in terms of product features and their respective sentiments has been the most appreciated form of fine-grained representation [21,22]. The

* Corresponding author.
E-mail addresses: v_manju@blr.amrita.edu (M. Venugopalan), g_deepa@blr.amrita.edu (D. Gupta).

[1] Sentiment Analysis turns Customer Reviews into Insights---Vendasta Blog.

https://doi.org/10.1016/j.knosys.2022.108668
0950-7051/

<text>Chow fun was dry; pizza was hard, pork shu mai was more than usually greasy and had to share a table with loud and rude family.
</text>
<Opinions>
*<Opinion polarity="**negative**" category="**FOOD#QUALITY**" target="**Chow fun**"/>*
*<Opinion polarity="**negative**" category="**FOOD#QUALITY**" target="**pizza**"/>*
*<Opinion polarity="**negative**" category="**FOOD#QUALITY**" target="**pork shu mai**"/>*
*<Opinion polarity="**negative**" category="**AMBIENCE#GENERAL**" target="**NULL**"/>*

Fig. 1. Sample labeled review sentence from SemEval 2015 dataset.

last decade has witnessed research in many challenging areas like deep learning architectures for sentiment analysis [23–25], sentiment classification for codemixed data [26], explorations in low resource languages [27,28] and so on. The popularity of the field of sentiment analysis, owing to its immense applications has witnessed the emergence of a multimodal dimension [29,30] that extends sentiment analysis to other modalities like audio and visual data beyond text. The emergence of fields like affective computing and sentiment analysis [31–33], which tries to explore the strength of combined analysis of sentiment and emotion has provided new dimensions to the field.

The early years of the decade of 2000 saw the emergence of a new field called aspect level sentiment analysis [34], which demanded a fine-grained sentiment representation in terms of the product features/aspects and their respective polarities. A user's affinity towards a product is often based on aspects of interest or priority rather than an overall rating. Aspect based sentiment analysis helps a customer to make a choice of a product/service based on the aspects of his preference. The same applies to a situation where a company needs to know the highly appreciated or underrated aspects of a product, to rectify the model. This elevates the importance and demand for aspect based sentiment analysis. For example, in the review sentence, *"The restaurant had a great ambience but food was pathetic"*, the aspect terms are '*ambience*' and *'food'*. The subtask which identifies the aspects is called aspect term extraction. This subtask is then extended by finding the polarity orientation towards each identified aspect which maps the polarity of *'ambience'* as *'positive'* and *'food'* as '*negative*' and is called sentiment classification. Fig. 1 depicts a sample labeled instance for aspect based sentiment analysis from benchmark SemEval datasets. The aspect terms are marked by the label *target.* The category to which aspect terms belong to and its polarity are marked by labels *category* and *polarity* respectively. As observed in Fig. 1, '*Chow Fun*' is an aspect term labeled by the tag *target*, its corresponding aspect *category* is labeled as '*FOOD#QUALITY*' and its *polarity* is marked *'negative'*. The targets in a review text could be single word aspects like *'pizza'* or multi-word aspects like *'pork shu mai'* or *'Chow fun'*. The subtask of aspect term extraction which identifies the aspect terms labeled as *target* is the focus of this work.

There has been active research carried out in the field of aspect term extraction [35,36], mostly supervised. Supervised models have been successful to an extent with respect to aspect term extraction, provided the size and the quality of the labeled data suffices. Often, supervised models are dependent on massive sets of manually labeled training data. The difficulty of building huge volumes of labeled data for this fine-grained subtask encourages the need for unsupervised models [37,38]. The data labeling industry is worth a billion-dollar business[2] in the field of AI and often struggles in ensuring the quality of data labels for fine-grained labeling tasks. Hence weakly supervised or unsupervised models are the need of the era. A few semi-supervised models have been experimented for aspect term extraction but have not produced satisfactory results. Aspect based sentiment analysis is crucial for decision making in many areas like product design, market research, customer support etc. Therefore, a compromise cannot be made on the model performance. Hence an effective unsupervised model is the goal.

The proposed model is almost unsupervised, based on guided Latent Dirichlet Allocation (LDA) [39]. LDA models experimented earlier for aspect term extraction lacked in their ability to differentiate the overlapping topics that represent each aspect category. Moreover, to target multi-word aspects, rather than incorporating all N-grams to the LDA model, they should be filtered carefully based on probable regular expressions representing multi-word aspect occurrences. The traditional LDA also fails to capture certain hidden semantic similarities. The proposed model overcomes these limitations. The contribution of the proposed approach can be briefed as

- It proposes a guided LDA model that uses only a few generic seed words from each aspect category combined with an automatic seed set expansion module based on BERT similarity resulting in better and faster topic convergence. The supervision required for the model is minimal seed aspect terms from each aspect category which can be known from generic knowledge of the domain.
- The inputs to the guided LDA model are crafted using regular expression (RE) based linguistic rules, which helps in better targeting multi-word aspects. These guided inputs are filtered using multiple pruning strategies that majorly includes a BERT based semantic filter to incorporate semantic strength in situations where co-occurrence statistics may fail to serve as a differentiator.
- The threshold parameters for seed set expansion and filtering of RE-based inputs have been tuned using Particle Swarm Optimization (PSO) strategy.

The organization of the paper comes in following order: Section 2 illustrates some of the prominent works related to aspect term extraction; Section 3 showcases the details of the proposed model, Section 4 describes the experimental setup in terms of corpus details, parameter tuning and evaluation metrics. Section 5 presents the experimental results and error analysis, and finally Section 6 presents the conclusions and future work directions.

2. Related works

Experimental research in aspect term extraction has been supervised or unsupervised in nature with the former being more explored. Minimal supervision being the demand of the time, the focus of the survey has been confined majorly to unsupervised or weakly supervised models for aspect term extraction. Under unsupervised category, frequency based approaches, syntax

[2] https://medium.com/syncedreview/data-annotation-the-billion-dollar-business-behind-ai-breakthroughs-d929b0a50d23.

based, topic modeling and hybrid deep learning models have been discussed. Few supervised deep learning models have also been surveyed.

The initial approach to aspect extraction identified nouns and noun phrases in the given text and applied frequency based filters. Frequency based approaches applying association rule mining [40,41] helped in determining frequent item-sets as product features. Such models have been enhanced using compactness and redundancy pruning strategies. Baseline statistics [42] to prune false positives and incorporating grammatical dependencies to identify infrequent aspects are a few among the improvisations experimented on frequency based models. Syntax based approaches [43,44] based on the underlying syntactical dependencies have been experimented to identify aspects. Frequency based filters and similarity-based approaches based on co-occurrence statistics have been used to improve model precision. The infrequent aspect coverage has been improved by measuring its relation to domain specific opinion words. Most of the early experimentations on aspect term extraction have been performed on the Hu and Liu 2004 dataset[3] which consists of customer reviews from the electronic domain.

Topic models are unsupervised models which unfold to discover hidden patterns in text. It tries to map the co-occurrence of words in different documents with their probability of belonging to the same topic. The concept of topic modeling when applied in Sentiment Analysis ends up in identifying the entities and attributes as topics in the text. Mei et al. applied a Probabilistic Latent Semantic Indexing based topic modeling approach [45] to extract aspects from text. The approach tried to jointly model topics and their associated sentiment and presented results in terms of a set of words representing each topic. The proposal of a new feature reduction technique called Latent Dirichlet Allocation (LDA) [46], threw open new arenas in topic modeling. Titov and McDonald presented their findings where they argued that topics produced by LDA do not map to the aspects of entities but are more associated with topics that identify tokens as entities. The multi-grain LDA model [47] proposed to overcome this disadvantage, modeled topics into global and local categories. The global topic distribution is assumed to be fixed for a document but local topic distribution can vary. Titov et al. also proposed a model [48] to discover aspects from reviews, reinforced by textual evidence from reviews. Brody et al. used a basic LDA implementation [49] to model aspects and its sentiments. To incline the model towards localized aspects rather than global topics in the text, each sentence is treated as a separate document and the distribution of aspects for each sentence is the expected output of the model. Another variation of LDA [50] eliminates the bag-of-words assumption and assumes that the topics of words in a document form a Markov chain, and that subsequent words are more likely to have the same topic. The model reported appreciable improvement in Rand Index measure when reported on Hu and Liu dataset in comparison to standard topic models.

In recent years, deep learning models have been proposed for aspect term extraction. The majority of supervised models [51–53] experimented are variations of Convolution Neural networks (CNN) architecture. Dependency-tree based convolutional stacked neural network, which captures syntactic features, controlled CNN for asynchronous parameter updation, enhancement using double embeddings, etc. are the variations attempted. These supervised models have reported appreciable improvement from the state of art on benchmark SemEval datasets. Akhtar et al. proposed a cascaded BiLSTM and CNN network [54] for joint learning of aspect and sentiment extraction. An unsupervised approach in this direction includes a neural auto encoder framework [55] where the sentence representation is enhanced by combining the word embeddings of relevant sememes which are basic units of morphemes. The proposers claimed that the approach helped in efficient extraction of latent semantic information from sentences. An attempt to incorporate background knowledge in an LSTM network by augmenting a stacked attention mechanism [56] achieved an improvement over the state of art. Hybrid two phase approaches [57–59] have used linguistic rules followed by a pruning stage based on domain correlation to extract candidate phrase chunks in phase1. These chunks are in turn used as pseudo labels to train a GRU or a Bi-LSM network for aspect term extraction in phase2. These hybrid approaches have been able to elevate the performance beyond the state of art unsupervised deep learning models. A challenge recently addressed in fine-grained sentiment analysis includes ambivalence handling [60,61] which deals with bi-polar comments towards an aspect. Integration of symbolic and sub-symbolic AI tools, which is an attempt towards more interpretable AI, has been applied in sentiment analysis to enhance polarity detection [62]. Neural tensor networks [63] are also being experimented in sentiment analysis [64,65] to extract contextual polarity and thus contribute in polarity detection.

Concerning the subtask of aspect term extraction, the frequency based approaches experimented initially were simple and straightforward but were not effective on benchmark datasets like SemEval, especially for multiword aspects. Rule based approaches which tried to capture the syntactic information were not able to generate competitive performance as stand-alone models. They are also constrained by the coverage of manually crafted rules and the accuracy of such models is challenged when the review sentences are grammatically incorrect. As far as topic modeling approaches are concerned they still need to be fine-tuned for applications like aspect level sentiment analysis where the topics are not very distinct. Deep Learning models have been rarely experimented in a purely unsupervised manner. Mostly the experimentations witnessed an unsupervised or semi-supervised model used to predict candidate aspects which are later used as pseudo labels for training a deep learning model.

As the challenges of data annotation in terms of cost and quality especially for such fine-grained tasks is a burden, more research needs to be explored on unsupervised or weakly supervised methods for aspect term extraction. The application being a crucial input to decision support systems in the industry, the performance metric also cannot be compromised. Based on literature survey, a topic modeling approach fine-tuned for the application of aspect term extraction is a promising approach in the unsupervised direction. Even though LDA is the state of art in topic modeling, in scenarios where the corpus is not characterized by discrete topics, the purely unsupervised nature of LDA fails to produce meaningful topics which is typically applicable for a problem like aspect term extraction. LDA is expected to group aspect terms into topics based on aspect categories. But since aspect categories themselves are not that distinct, LDA would turn out to be weak in converging to the expected topics. Hence a guided LDA model which takes inputs in the form of minimal aspect seed words corresponding to aspect categories would be guided to learn topics of specific interest to the user. Such a framework would be an ideal choice for the proposed model. Hence the model proposed for aspect term extraction is a guided LDA model enhanced by automatic aspect seed set expansion, guided inputs and multiple pruning strategies which include Named Entity Recognition (NER) and frequency based filters and a BERT based semantic filter which provides an additional layer of semantic strength.

3 https://www.cs.uic.edu/~liub/FBS/sentiment-analysis.html.

3. Proposed methodology

The work aims to design an unsupervised approach for aspect term extraction from product reviews using a guided LDA model, which is the backbone of the proposed approach. The guided inputs to the guided LDA model are in two forms. The N-grams filtered by regular expressions and thereafter subjected to multiple filtering stages form the first input. The second input is in the form of a seed set representative of aspect categories. The schematic diagram of the enhanced guided LDA model for aspect term extraction is depicted in Fig. 2 and is explained in detail in the following subsections. The methodology commences with a text pre-processing module in stage 1 followed by an input sequence generation and filtration module which is stage 2. As observed in Fig. 2, stage 3 depicts the aspect seed set expansion module and is followed by the Enhanced Guided LDA module represented as stage 4.

3.1. Text pre-processing module

This module (depicted as stage 1 in Fig. 2) incorporates the generic pre-processing techniques on textual data specifically fine tuned to the application handled. In this pre-processing phase, the raw opinionated sentences are extracted from the unlabeled raw training data of the dataset (shown as Input-1 in Fig. 2) and it goes through stages like removal of special characters and punctuations which do not contribute to opinion. A decision was taken not to remove all the punctuations bluntly. For example, the sentence *"This is one of my favorite spot, very relaxing the food is great all the times, celebrated my engagement and my wedding here, it was very well organized."* has three clauses just separated by commas and its POS tagging quality would be compromised by removing the comma separators. Lower case conversion was not performed as it would affect the quality of POS tagging and NER extraction in certain cases. For example, in the sentence "*The restaurant is located in New York.*" the word *new* would be tagged as an adjective if it is in lower case and *'New York*' would not be identified as an entity by the NER extractor. Replacing contractions like *I'd* with *'I had', she's* with *'she is', 'can't' with 'cannot'* etc. is performed as it improves tagging quality. After noise removal, the inputs are subjected to tokenization, POS tagging and NER tags extraction. This pre-processed data is then input to the *Input Sequence Generation and Filtration* module depicted as stage 2 in Fig. 2.

3.2. Input sequence generation and filtration module

This module deals with the generation of input sequences from pre-processed data, based on a set of regular expressions derived from linguistic rules. These input sequences are then subjected to various filtration stages, which are explained in detail in the consequent subsections. The different stages in the module are showcased in Fig. 2.

3.2.1. Generating input word/word sequences based on regular expressions

From the pre-processed data, input sequences are generated based on a set of regular expressions listed in Table 1, which reflect prospective single and multi-word aspect terms. In general, nouns and noun phrases in review sentences are more likely to be aspects. Grounded on that fact, the regular expressions listed in Table 1 are crafted from highly reliable rules based on knowledge derived from our prior rule based implementation [37] and other rule based approaches [44,66]. A set of minimal rules with comparatively better reliability were chosen that required only minimal linguistic resources (POS tags, NER tags and an opinion lexicon) to reduce cumulative errors. The regular expression in Table 1, "N*" indicates a sequence of nouns, "JN+" indicates an adjective followed by at least one noun and so on. The "JN+" regular expression is subjected to another condition that the token corresponding to the adjective should not be an opinion word. This condition is implemented using Bing Liu opinion lexicon.[4] Thereby each review sentence is converted into a sequence of N-grams that adhere to the regular expressions in Table 1. In order to capture the association between aspect words and opinion words, adjectives in the review sentence are also appended to the input sequence of each review sentence. All single word tokens in the input sequence corresponding to the regular expression "N" are also subjected to lemmatization. It was observed that single word aspects occur with almost equal weightage in singular and plural forms e.g. "*price/prices", "dish/dishes", "portion/portions"* and so on. This elevates the importance of performing lemmatization at this stage. Let this input sequence be depicted as $Input_{RE}$.

3.2.2. NER, semantic and frequency based filtration of input sequences

The input sequence $Input_{RE}$ is subjected to NER, semantic and frequency based filters which passes through three stages and is depicted in Algorithm 1.

STAGE 1: NER based filtering

A mechanism to filter false positives, especially multi-words aspect terms using NER filters, is incorporated in STAGE 1 of the algorithm. Word sequences like "*saturday night*" or "*late afternoon*" which correspond to the regular expressions "N*" and "JN+" respectively are wrongly extracted as candidate aspect terms. Such wrong entries can be identified by using a few NER labels like 'PERCENT', 'LANGUAGE', 'MONEY', 'TIME', 'DATE', 'ORDINAL', 'QUANTITY', 'CARDINAL', 'DURATION', 'SET' and 'NUMBER' from the raw text entries and such wrong candidates can be removed from the input sequence. A few examples of effective filtering enabled through NER filtering is presented in Table 2. $Input_{RE}$ after being subjected to NER filtering is passed on to STAGE 2 of Algorithm 1.

STAGE 2: Filtering high frequency extremes

Filtering N-grams with very high/low frequency is a usual approach in topic modeling approaches as they are not expected to contribute much to topic clusters. With respect to the application handled, filtering very high frequency words like '*restaurant*' which are characteristic words in the domain would be beneficial. Hence, the frequency of all the unique input sequences in the corpora is found and very high frequency outliers that are domain representatives are removed, represented as STAGE 2 in Algorithm 1. The filtered input sequence is passed on to STAGE 3.

STAGE 3: Semantic filter implementation in conjunction with low frequency filter

A decision similar to that of high frequency filter cannot be implemented for the lower frequency filter as the frequency of occurrence of a prospective aspect term could be as low as just once. Hence a pure frequency based filter cannot be applied for the lower bound. A BERT [67] similarity based semantic filter which guarantees context specific similarity (when compared to the prior word embedding models) is appended to the lower bound frequency filter. The lower frequency bound is applied only after approval by a semantic filter $\theta_{sem-low}$ which stands for the minimal semantic similarity an input token should have with any one word in the aspect seed set. BERT based semantic similarity for a word-pair is measured in a context sensitive

[4] https://www.cs.uic.edu/~liub/FBS/sentiment-analysis.html#lexicon.

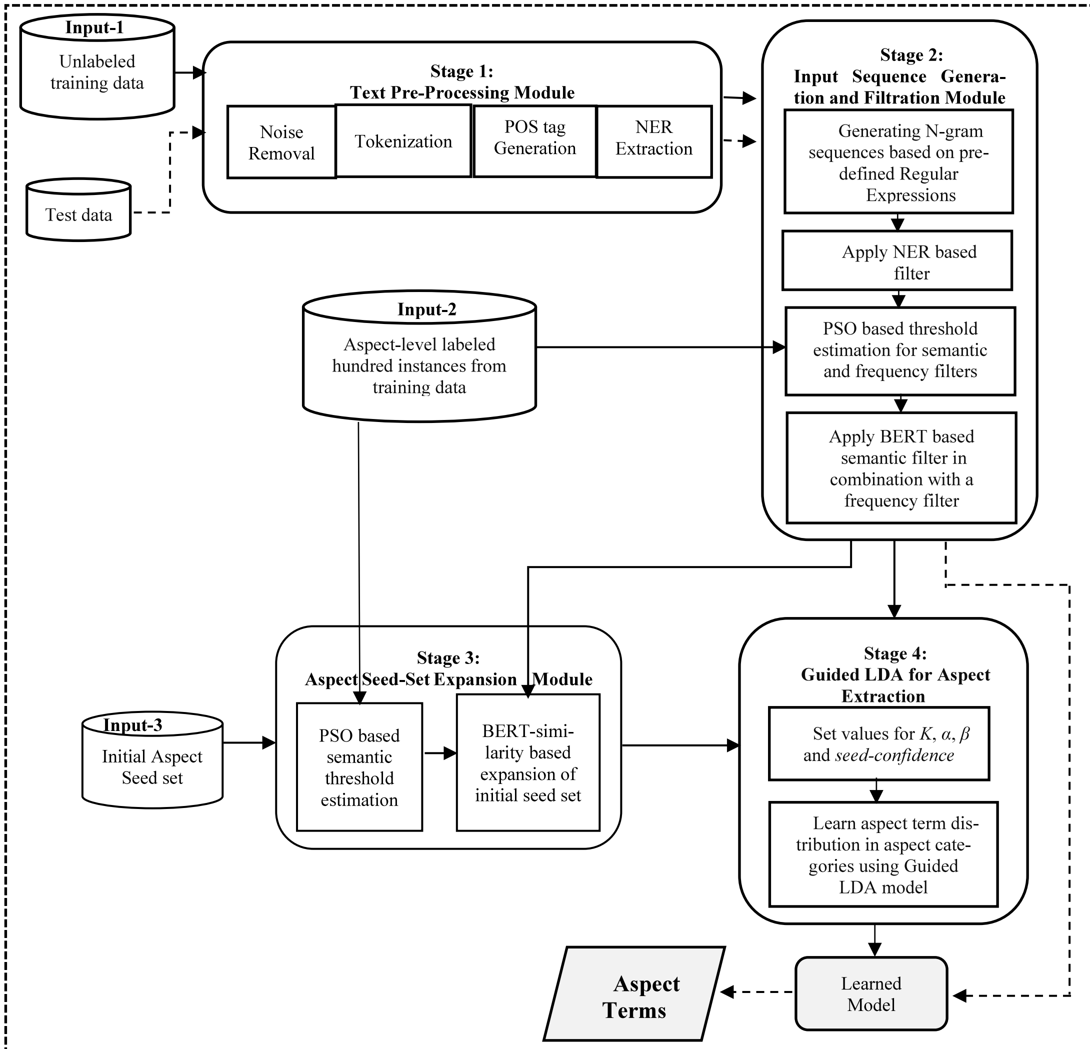

Fig. 2. Schematic diagram of the enhanced guided LDA model for aspect term extraction. (The stages marked indicate the process flow. The train data and test data flow are depicted using solid and dashed lines respectively.)

Table 1
Regular expressions for prospective aspect terms.

Regular expression	Comments	Examples
N	All single word nouns	Food, sushi, salads
N*	Sequence of nouns	Staff members, delivery service
J N+	Adjective followed by nouns where the adjective is not an opinion word	Fresh juices, huge portions, fresh leaf salad
N* IN N*	Nouns with a preposition in between	Glass of wine, pizza with soy cheese
N* IN DT N*	Nouns with a preposition and determinant in between	Tart of the day, portions of the food
VBN/VBP N*	Past or past participle form of verb followed by noun sequences	Roasted calamari, fried potatoes

Table 2
NER based filtering.

Term pruned	Regular expression	NER tag
Twenty minutes	JN+	TIME
Past summer	JN+	DATE
Saturday night	N*	TIME
New York	N*	GPE
New Years Eve	N*	EVENT
Last weekend	JN+	DATE
Restaurant week	N*	DATE

manner. The vector representation for a word is retrieved by inputting the sentence which represents the word context. Later the semantic similarity between a word pair is measured as the cosine similarity of the vector pair. Optimal values for both the thresholds θ_{f-low} and $\theta_{sem-low}$ are estimated using the PSO approach, a popularly used technique for parameter tuning in various domains [68].

With respect to PSO based estimation, the idea is to find optimal values for the parameters $\theta_{sem-low}$ and θ_{f-low} by using a randomly chosen aspect level labeled dataset of 100 reviews from the training data referred to as *Input-2* in Fig. 2. This is based on the belief that word embedding based semantic similarities follow a normal distribution [69] are almost uniform across different random samples. The parameter $\theta_{sem-low}$ is the low semantic similarity bound used in combination with a low frequency bound to remove false positives. But the right choice of both these parameters should ensure a balance between precision and recall of aspect terms and hence the fitness function for

Algorithm 1: NER, BERT-semantic and frequency based filtration of Input Sequences

Input: The input sequences based on regular expressions*($Input_{RE}$)*, a basic set of 15 seed words uniformly distributed across five aspect categories *(S)*, a list of noun chunks identified from raw unlabeled input training text using non-aspect NER labels *(NER)*
Output: The filtered input sequence ($Input_{FRE}$)

```
##STAGE 1 filter prospective aspects using NER filters
for i in range(len(Input_RE)):
    for j in range(len(Input_RE [i])):
        if Input_RE [i][j] in NER :
            remove Input_RE[i][j] from Input_RE[i]
    end
end

##STAGE 2 filter high frequency extremes
for i in range (len(Input_RE)):
    for j in range (len(Input_RE[i])):
        if frequency(Input_RE[i][j]) is a high frequency outlier:
            remove Input_RE[i][j] from Input_RE [i]
    end
end

####STAGE 3A create a list of N-grams satisfying minimal semantic threshold
Estimate θ_sem-low and θ_f-low using PSO
pass=S
Create a unique list V from Input_RE
while (1):
    L=len(pass)
    for i in range(len(V)):
        max=0
        for j in range(len(S)):
            sim=COSINE_SIM(BERT_VEC(V[i]),BERT_VEC(S[j] )):
            if sim>max:
                max = sim
        end
        if max> θ_sem-low:
                append V[i] to pass
    end
    L1=len(pass)
    if L1==L:
        break
end

##STAGE 3B apply low frequency filter in conjunction with a semantic filter
for i in range (len(Input_RE)):
    for j in range(len(Input_RE [i])):
        if frequency(Input_RE [i][j])<= θ_f-low and Input_RE [i][j] not in pass :
            remove Input_RE [i][j] from Input_RE [i]
    end
end
Input_FRE= Input_RE
```

PSO is defined as Specificity + Sensitivity-1, where *Specificity* and *Sensitivity* correspond to when $\theta_{sem-low}$ and θ_{f-low} are chosen as the semantic and frequency thresholds for filtration in the proposed methodology i.e. in the implementation of STAGE 3 of Algorithm 1. The fitness function is normalized in the range 0 to 1.

Each particle in the initial population of the PSO environment would be of the form [$\theta_{sem-low}$, θ_{f-low}]. $\theta_{sem-low}$ is a real value in the range [0, 1]. θ_{f-low} is an integer bounded between [1, 10]. Particle position and velocity updates are done using standard PSO formulations. The PSO parameters are set and the algorithm iterates until convergence. The optimized threshold values are used to implement the lower frequency filter in combination with the semantic filter. As depicted in STAGE 3A of Algorithm 1, a list of unigrams and bi-grams in the input sequence with a semantic similarity of at least $\theta_{sem-low}$ with any of the seed words is generated. Finally, STAGE 3B applies the lower bound by removing N-grams with frequency less than θ_{f-low}, if the N-gram does not meet the minimum semantic threshold $\theta_{sem-low}$. This final filtered input is represented as $Input_{FRE}$. This filtered regular expression sequence $Input_{FRE}$ is input to the enhanced guided LDA model and is also utilized for seed set expansion as shown clearly in Fig. 2.

3.3. Aspect seed set expansion module

A seed set that is representative of the aspect categories should be input to the guided LDA algorithm. The dataset in consideration has aspect terms from the restaurant domain broadly belonging to five aspect categories *Food, Ambience, Price, Service and General* category, which are representative of the topics. A set of any three commonly observed aspect terms in each of the five aspect categories is manually chosen to form a set of 15 seed words *S*, which are presented in Table 3. These seed words are represented as *Input-3* in Fig. 2 and are derived from generic knowledge of the domain. An automated procedure is used to expand the initial seed set, which is explained in Algorithm 2.

Table 3
Set of seed words for restaurant domain.

Aspect category	Seed words
Food	*Meal, chicken, drinks*
Ambience	*Atmosphere, ambience, décor*
Service	*Service, staff, delivery*
Price	*Price, cost, money*
General	*Experience, place, crowd*

An iterative procedure examines each unigram and bi-gram in the filtered input sequence $Input_{FRE}$.Any N-gram with a BERT based word embedding similarity value greater than $\theta_{seed-sem}$ with any of the seed set tokens is added to the seed set. This sub-procedure is iterated until the seed set stops expanding as depicted in line 16 of Algorithm 2. The goal is to expand the seed set with more seed words in each category to the extent possible but with a complete surety, hence the threshold $\theta_{seed-sem}$ is set to a high similarity score derived through a PSO based threshold estimation. This iterative procedure for automatic seed set expansion contributes to a significant enlargement of the seed set and would also ensure that choice of seed words would not create a noticeable impact on the Guided LDA results.

Algorithm 2: Aspect Seed Set Expansion

Input: $Input_{FRE}$ the filtered input sequences, S the seed set
Output: The expanded seed set S_{exp}

```
#### expand the initial seed set
1.  Estimate the value of θ_seed-sem using PSO
2.  Create the unique list of unigrams and bi-grams V from Input_RE
3.  while(1):
4.      L=len(S)
5.      for i in range(len(V)):
6.          max=0
7.          for j in range(len(S)):
8.              sim=COSINE_SIM(BERT_VEC(V[i]),BERT_VEC(S[j])):
9.              if sim>max:
10.                 max= sim
11.         end
12.         if max> θ_seed-sem:
13.             append V[i] to S
14.     end
15.     L1=len(S)
16.     if L1==L:
17.         break
18. end
19. S_exp= S
```

The optimal value for the parameter $\theta_{seed-sem}$ is found by using a randomly chosen aspect level labeled dataset of 100 reviews from the train data (depicted as Input-2 in Fig. 2). The parameter $\theta_{seed-sem}$ is a high similarity bound used to expand the seed set where there should be complete surety on the selection of the seed set which would imply that the fitness function for PSO would be minimization of False Positive Rate. Hence each particle in the initial population would be of a single dimension of the form $[\theta_{seed-sem}]$ which would take a value between 0 and 1 as the BERT similarity values are in the range [0, 1]. The minimization of false positive rate tending to zero is the straightforward visible objective. But that would optimize $\theta_{seed-sem}$ to a value 1, which would not leave scope for seed set expansion at all. Hence the fitness function is defined as the summation of two measures, *Penalty* and *FPR*. The measure *Penalty* is a penalty imposed as the semantic similarity threshold increases beyond 0.5 which is defined by Eq. (1). *FPR* is the false positive rate when $\theta_{seed-sem}$ is used as the semantic threshold parameter in the implementation of Algorithm 2. The particle position and velocity updates are done using standard PSO formulations. The PSO parameters are set and the algorithm iterates until convergence.

$$Penalty = \frac{\theta_{seed-sem} - 0.5}{0.5} \tag{1}$$

3.4. Enhanced guided LDA for aspect term extraction

The backbone of the proposed approach for aspect term extraction is a guided LDA model enhanced by guided inputs and the expanded aspect seed set. The filtered input sequence $Input_{FRE}$ (from stage 2 in Fig. 2), which is extracted and filtered from the review sentences of the training data forms the first input to the guided LDA. The expanded aspect seed set S_{exp} (from stage 3 in Fig. 2) which is representative of the aspect categories in the corpora forms the second input and would guide the model to improve both topic-word distributions and review-topic distributions. These two inputs are fed to the Guided LDA model as observed in Fig. 2. The plate notation that depicts the enhanced guided LDA model is presented in Fig. 3. The parameter K which implies the number of topics is mapped to the number of aspect categories with respect to aspect term extraction. The values for guided LDA hyperparameters α, β and *seed-confidence (SC)* which is the reliability factor of seed words are set for the Guided LDA algorithm. The generative model for guided LDA is depicted in Algorithm 3.

Each of the K topics which map to aspect categories is defined as a mixture of two multinomial distributions, a seed topic distribution ϕ^{Sexp} and a regular topic distribution ϕ^{r}. They are chosen using Dirichlet priors via the β parameter as shown in lines 1–4 of Algorithm 3. The seeded version ϕ^{Sexp} ensures that topics are guided by seed words from the expanded seed set S_{exp}, while the regular version ϕ^{r} allows other relevant dimensions to emerge. In the direction of improving review-topic distributions using seed words, each of the K seed sets is associated with a multinomial distribution called the group-topic distribution ψ, which are generated using Dirichlet priors controlled by the hyper parameter α as shown in lines 5–7. These group-topic distributions are used as prior to draw the review topic distribution θ. For each review sentence, a binary vector b of size K is defined (line 9) indicating which aspect seeds exist in the review. For example, if the input sequence corresponding to a review sentence is ['*meal*', '*crowd*'] the b vector would be defined as [1, 0, 0, 0, 1] that defines a means of a Dirichlet distribution from which a review-group distribution ζ_r is sampled. The concentration of this Dirichlet is set to a hyperparameter τ, thus, $\zeta_r \sim Dir(\tau^b)$ as shown in line 10 of the algorithm. From the resulting multinomial, a group variable g is drawn for this review sentence. This group variable brings a clustering structure among the reviews by grouping the reviews that are likely to talk about the same seed set. Once the group variable g is drawn, the review-topic distribution θ is chosen from a Dirichlet distribution with the group's topic distribution as the prior, which ensures that the topic distributions of reviews within each group are related. The first iteration starts by assigning each word in R review sentences to a topic, where the topic is chosen based on the value of x. x is a binary parameter that chooses between a seed distribution and a regular distribution based on the value of a randomly generated variable and *seed confidence* (SC) as observed in Fig. 3. A topic z among K topics is chosen for every token w among the N_r tokens in the review sentence derived from either ϕ^{Sexp} or ϕ^{r}. For example, if the current token w is '*chicken*' and if the variable x is favorable for a seed topic allocation then the topic assigned to the token will be *Topic1* (as per information in Table 3). The expanded seed set S_{exp}, thus ensures a high probability that words which belong to the same aspect category are assigned to the same topic by the guided LDA model.

Algorithm 3: Enhanced Guided LDA

Input: $Input_{FRE}$ the filtered input sequences, the expanded seed set S_{exp}, *K number of aspect categories*, α, β and *seed-confidence (SC)*
Output: Probability of each input token belonging to each of the K topics

1. for t in range(K):
2. Select ϕ^{r}_{t}~Dirichlet(β).
3. Select ϕ^{Sexp}_{t}~Dirichlet(β)
4. *end*
5. for s in range (K):
6. Select a group-topic distribution $\psi_s \sim$ Dirichlet(α)
7. *end*
8. Repeat until convergence:
9. for r in range (R):
10. Select a binary vector b of length K
11. Select a review-group distribution $\zeta_r \sim$ Dirichlet(τ^b).
12. Select a group variable g $\sim$ Multinomial(ζ_r)
13. Select $\theta_r \sim$ Dir(ψ_g)
14. for each token i in N_r:
15. Select a topic $z_i \sim$ Multinomial(θ_r)
16. Generate a value for x using a random generator and SC
17. if x is 0:
18. Select a word $w_i \sim$ Multinomial(ϕ^r)
19. if x is 1:
20. Select a word $w_i \sim$ Multinomial(ϕ^{Sexp})
21. *end*
22. *end*
23. *end*

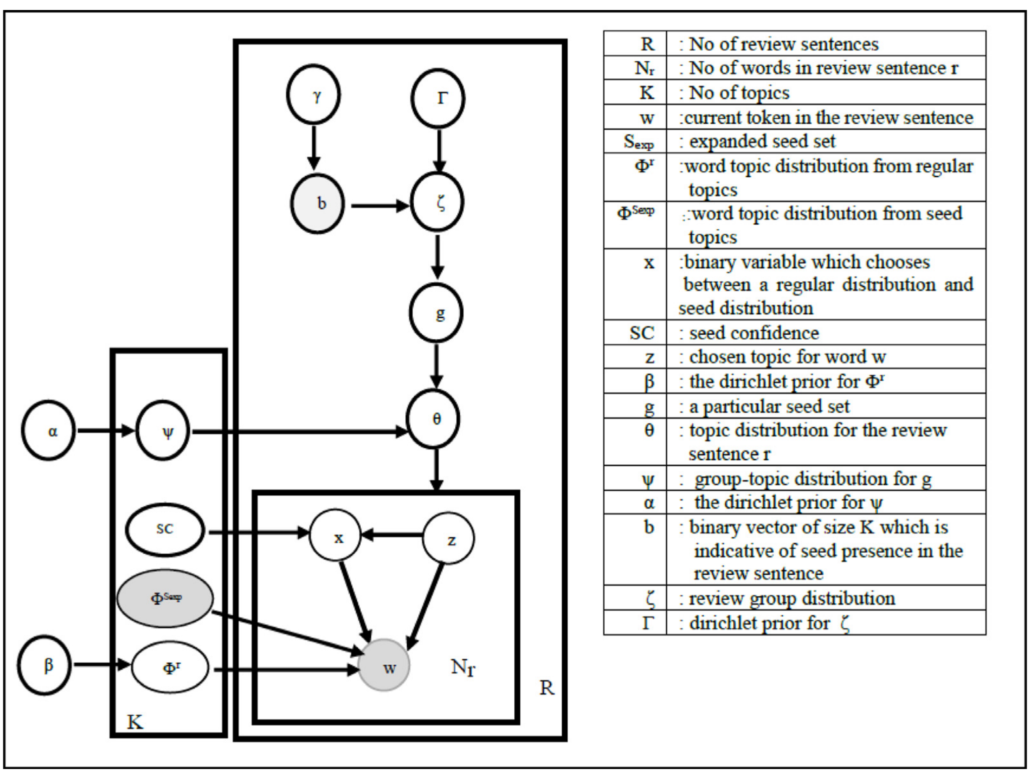

Fig. 3. Plate notation for enhanced guided LDA.

In every further iteration, the words in the review sentences get assigned to a different topic so as to maximize the product of review-topic affinity and topic-word affinity. One iteration is sketched in line 9–22 of Algorithm 3. The algorithm runs iteratively until convergence. The final output of the guided LDA algorithm is the probability of each token in the input belonging to each of the K topics. The first level of enhancement of the proposed architecture beyond the basic guided LDA model is brought in by the filtered input sequence represented as the tokens w of R review sentences in Fig. 3. The expanded seed set provides the second layer of enhancement by ensuring that the probability of two similar words to be categorized in the same topic is boosted in the proposed guided LDA architecture based on its co-occurrence similarity with aspect seeds in S_{exp} representing that topic.

Table 4
SemEval data statistics for the restaurant domain.

Year	# of sentences in train data	# of sentences in test data	# of single word aspects in test data	# of multiword aspects in test data
2014	3041	800	818	316
2015	1315	685	339	152
2016	2000	676	485	165

3.5. Extracting aspect terms from test data

The input sequences of the test data pass through the text pre-processing module and the multiple stages of the filtration module. Finally, when fed to the learned enhanced guided LDA model, the model predicts the candidate aspect term distributions among the aspect categories. The probability of every token in the test data belonging to each of the topics/aspect categories is extracted. The tokens that have a higher probability of belonging to each aspect category are believed to be aspect terms. The candidate aspects corresponding to the lowest probability score in each aspect category are discarded and the rest are considered as the final aspect terms. Thus single and multi-word aspect terms are extracted from the input test data using the proposed enhanced guided LDA model.

4. Experimental setup

4.1. Corpus details

The proposed approach has been experimented and evaluated on SemEval 2014, 2015 and 2016 datasets which are benchmark datasets for aspect term extraction. SemEval is a series of semantic evaluation workshops organized since 2010 and one of the tasks introduced since 2014 was aspect based sentiment analysis [70–72]. The dataset released by SemEval for the subtask of aspect term extraction contains data from restaurant and laptop domains. The proposed model has been experimented on the restaurant domain datasets of all three years. In 2014 datasets, the review sentences are annotated using aspect term and their polarities. 2015 and 2016 datasets on the other hand have labels for aspect terms, aspect categories and their polarities. A typical example of a review sentence from SemEval 2015 dataset is shown in Fig. 1. The proposed model is expected to predict the aspect terms listed under the tag *target.* The statistics of the train and test dataset are presented in Table 4. The train and the test data are characterized by single and multi-word aspects. Many review sentences do not contain any aspect terms. The test data of SemEval 2014 contains 194 review sentences out of 800 that do not contain any aspect terms. Despite a smaller size, SemEval 2015 test data contains 284 review sentences without any aspect terms and this count is 256 in SemEval 2016.

4.2. Hyperparameter settings and evaluation metrics

The various parameter and hyperparameter settings in the proposed model are discussed and the evaluation metrics are briefed. The proposed model has been implemented in Python and packages like nltk, numpy, sklearn, guidedlda etc. have been utilized. For POS tagging and NER extraction, Stanford Core NLP V4.0.0[5] has been used. The BERT model utilized is pre-trained BERT BASE. The Pytorch-Transformer library has been used to implement the BERT model.

The PSO based estimations of semantic and frequency thresholds mentioned in Sections 3.3 and 3.2.2 were carried out on a random sample of 100 aspect level labeled review sentences. This is based on the belief that distributions of word embedding based semantic similarities follow a normal distribution and are almost uniform across different random samples. A cross verification of this fact was performed by randomly selecting 10 different samples of 100 reviews each from train data. These were subjected to tokenization and stop word removal and their similarity with the initial seed set of 15 words was calculated. The mean and standard deviation of these values for each of the 10 samples considered is plotted in Fig. 4. As observed, the distributions are almost uniform and justify the fact that semantic thresholds can be determined by optimization techniques applied on any random labeled sample of just 100 reviews.

The optimal values of parameters $\theta_{sem-low}$, $\theta_{seed-sem}$ and θ_{f-low} are determined based on the experimental results of PSO performed on a random sample of 100 labeled instances(depicted as Input-2 in Fig. 2). The number of particles in the initial population is set to 15 and the algorithm is iterated 30 times which was suitable for convergence. Fig. 5 depicts the variation of the objective function with the threshold parameter $\theta_{seed-sem}$ on this random sample. It can be observed that the optimal $\theta_{seed-sem}$ value corresponding to the minimal value of the objective function is approximately 0.7. Similarly, Fig. 6 is a scatter plot which depicts the variation of the objective function with the threshold parameters $\theta_{sem-low}$ and θ_{f-low}. The maximum value of the objective function on the scatter plot corresponds to 0.4502 for $\theta_{sem-low}$ and 1 for θ_{f-low}. Thus the optimal values for $\theta_{sem-low}$, $\theta_{seed-sem}$ and θ_{f-low} are optimized to 0.45, 0.7 and 1 approximately. The optimized value of $\theta_{seed-sem}$ when used in the aspect seed set expansion module witnessed a considerable expansion of seed words from an initial count of 15 to almost 8 times its size when applied on the training data. This BERT-similarity based expansion module hence plays a major role in enhancing the performance of the guided LDA model.

With respect to the guided LDA parameters, the value of K, the number of topics is set to 5 as the broad aspect categories reflected across all the SemEval datasets are *Food, Ambience, Service, Cost and General* categories. The hyperparameters of the guided LDA model α and β which influence prior distributions of review-topic and topic-word probabilities respectively are set to the default values of 0.1 and 0.01 in the python package implementation of LDA. A low α value puts less constraints on review sentences and means that it is more likely that a review sentence may contain a mixture of just a few, or even only one of the categories. In other words, it would be a skewed distribution. Likewise, a low value for β means that each word would have a high probability to belong to one or lesser number of topics or aspect categories. This scenario seems to be suitable for the task under consideration and justifies the choice of α and β values. A third parameter of the Guided LDA algorithm is the *seed-confidence* value which shows the bias towards the seed set. It takes a value between 0 and 1. If the *seed confidence* value is set to 0.5, the seeded words are biased 50% more towards the seeded topics. The value of the *seed confidence* parameter has been set to 0.8 as the expanded seed set has been generated through a PSO strategy where the objective function was inclining towards zero false positive rate and hence the reliability on the aspect seed set is quite high.

Topic models being unsupervised, they would better learn distributions from larger volumes of unsupervised data. Hence for a better learning outcome, the raw text from train data of all three year datasets from SemEval (2014, 2015 and 2016) is input to the enhanced guided LDA model. Guided LDA clusters the input tokens into five clusters corresponding to aspect categories and

5 https://nlp.stanford.edu/software/lex-parser.shtml.

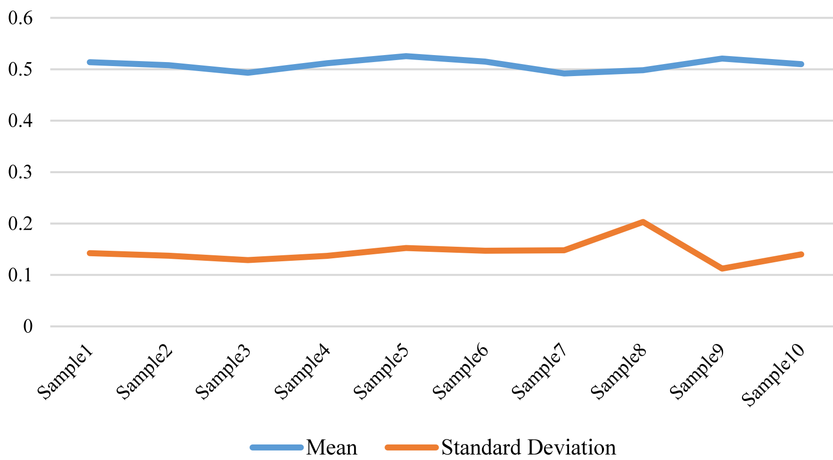

Fig. 4. Sample statistics of random samples chosen for PSO.

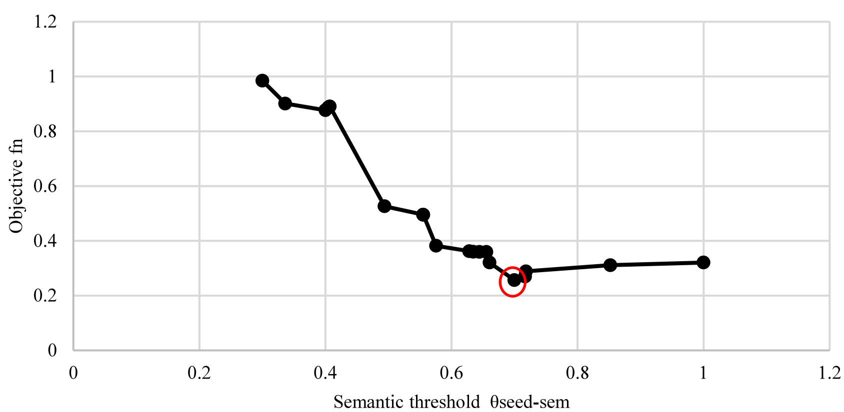

Fig. 5. PSO based estimation for $\theta_{seed-sem}$.

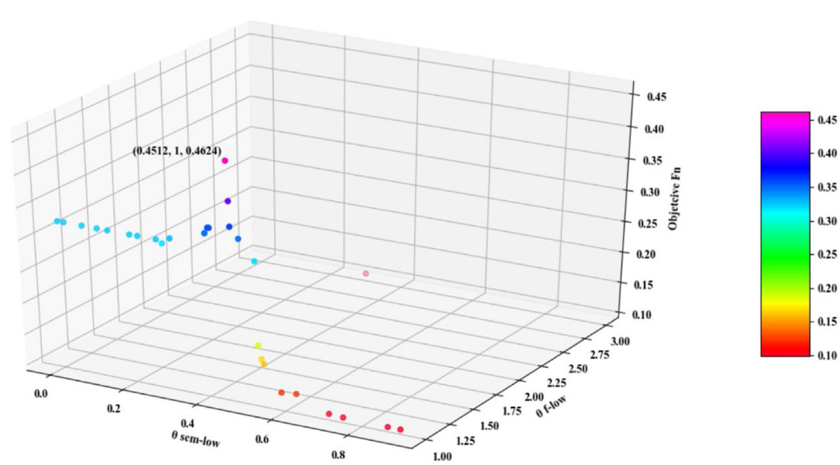

Fig. 6. PSO based estimation for $\theta_{sem-low}$ and θ_{f-low}.

maps a probability score of the token belonging to that category. The tokens are arranged in layers of descending probabilities. A token in the test data sequence is predicted as an aspect term if it belongs to a strata other than the lowest in any one of the five clusters. The proposed model performance has been evaluated on the test data sets of SemEval 2014, 2015 and 2016 datasets in terms of the standard metrics Precision, Recall and F-measure of the aspect terms. As far as multi-word aspects are concerned only completely recalled aspect terms are given credits by the evaluation system.

5. Results and analysis

The experimental results and analysis are presented under the following perspectives.

- The experimental results of the proposed methodology on SemEval 2014, 2015 and 2016 restaurant domain datasets in terms of the evaluated metrics are presented.
- The performance of the proposed model is also compared with the existing state of art baselines.

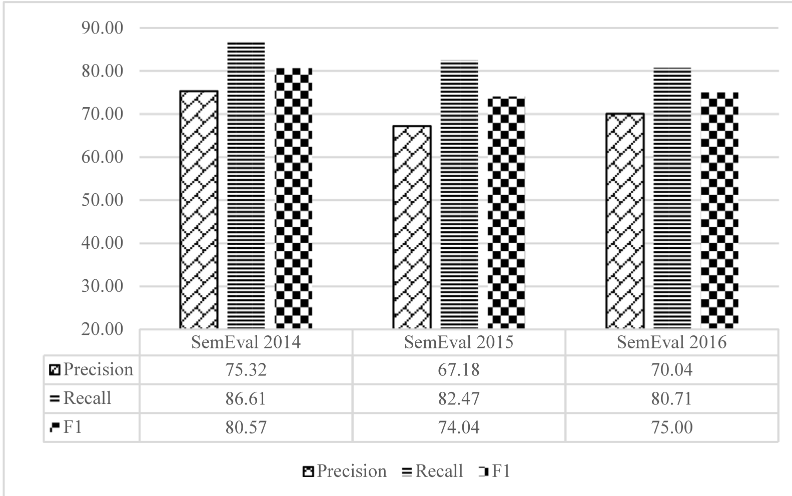

Fig. 7. Results of the proposed model for Aspect term extraction on SemEval datasets.

- A discussion about the factors that have negatively contributed towards the model performance.

5.1. Experimental results

The proposed methodology has been experimented on the Restaurant domain of SemEval 2014, 2015 and 2016 datasets and the results have been reported in Fig. 7 in terms of precision, recall and F-measure. The F-measure reported by the proposed model on SemEval 2014 dataset is almost 81% and is around 6 points higher when compared to the later year datasets. This is quite obvious from the fact that being a world class evaluation series the dataset would have been more challenging in the later years. As expected and reported by prior models, single words aspects are comparatively retrieved in larger volumes by the proposed model as revealed by the recall results displayed in Fig. 8. The recall of single word aspects in SemEval 2014 dataset is high at 94% and even in SemEval 2015 and 2016 datasets it is close to a 90% mark. The performance of the proposed approach on two word aspects has been quite appreciable achieving a recall of approximately 72%, 74% and 76% as observed in Fig. 8. Multi-word aspects with three and more words have been most challenging with minimum recall reported across all datasets, a 49% and 47% recall in 2014 and 2015 datasets and a further dip to around 34% in SemEval 2016. The results of the proposed model reported in Fig. 8 are with respect to complete recall of the aspect terms.

Fig. 9 similarly presents the precision achieved by the proposed model across single and multi-word aspect terms. It can be observed that a consistent and appreciable precision is achieved by the proposed model across single word aspects of all datasets. The highly balanced precision and recall of single word aspect terms as noticed in Figs. 8 and 9 to a large extent contribute to the F-measure achieved by the model. The considerable dip in the precision of multi-word aspects in comparison to the single word aspect terms is observed consistently across all three datasets. This would imply that the F-measure of the proposed approach is considerably impacted negatively by the false positives produced in the multi-word aspect category, as observed by the low precision in the range 40%–43% for multi-word aspects, across datasets except for two-word aspects in SemEval 2014 which is comparatively better at 54%.

5.2. Comparison of the proposed model with baselines

The proposed model performance has been compared with a few recent state of art unsupervised and supervised baselines which have reported results on SemEval datasets for aspect term extraction.

Unsupervised baselines

- Wu Chuhan et al. 2018 [57] is a two-step hybrid approach that uses chunk level linguistic rules in conjunction with pruning methods based on domain correlation to predict the aspect terms in phase 1 which have been used as pseudo labels to train a GRU network in phase 2.
- Dragoni et al. 2019 [73] proposed an opinion monitoring service implementing (i) a set of unsupervised strategies for aspect based opinion mining together with (ii) a monitoring tool supporting users in visualizing analyzed data. The aspect extraction strategies are based on the use of an open information extraction strategy.
- Chauhan et al. 2020 [58] combine linguistic patterns with a deep learning technique. The model uses rules-based methods to extract candidate aspects and targets domain-specific aspects using word embedding models. Finally, the extracted aspects are used as data labels to train the attention-based deep learning model for aspect-term extraction.

Supervised baselines

- Chen et al. 2017 [74] proposed a model which first applies a neural network model to classify opinionated sentences into three based on the number of targets that appear in a sentence. Each group of sentences is then fed into a BiLSTM based network for aspect term extraction.
- Hu et al. 2018 [53] propose a novel but simple CNN model employing two types of pre-trained embeddings for aspect extraction: general-purpose embeddings and domain-specific embeddings. They claimed to be the first attempt to make use of double embeddings.
- Wei Xue et al. 2017 [75] is a supervised classification method where aspect extraction is designed as a sequence labeling problem. The architecture comprises a Bi-LSTM network where a word embedding layer transforms the input words to real valued vectors.

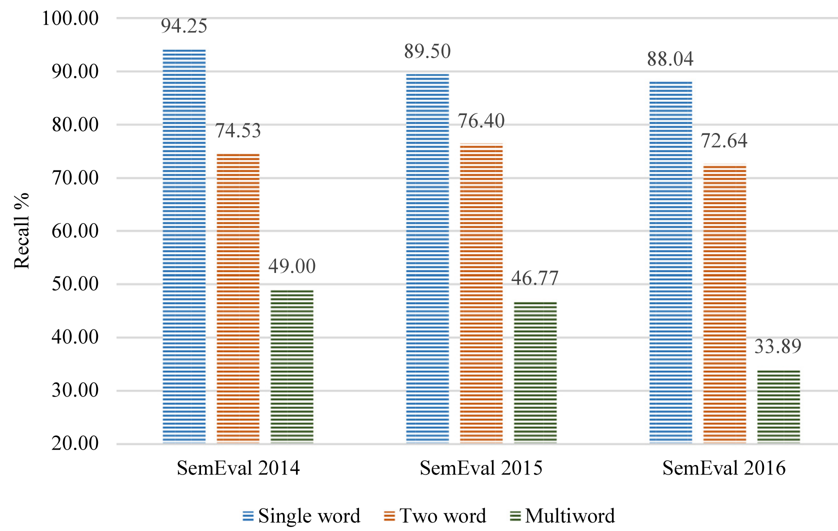

Fig. 8. Recall analysis of the proposed model across single and multi-word aspect terms.

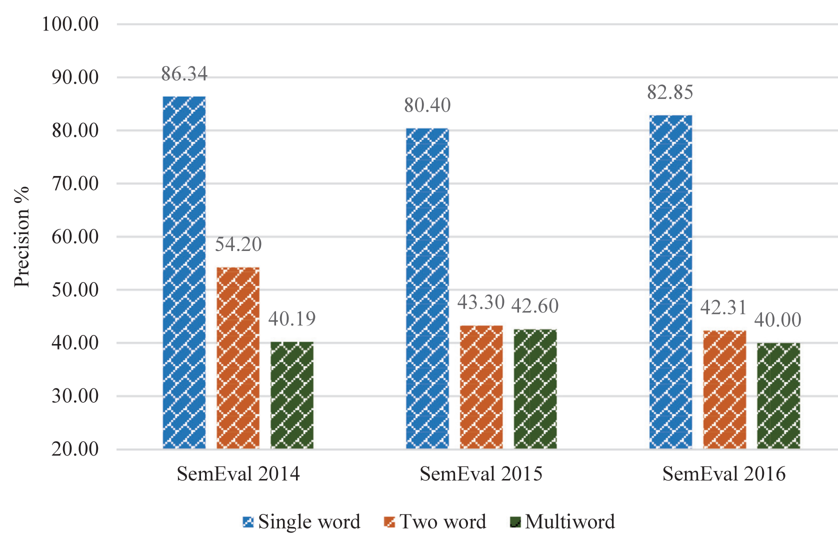

Fig. 9. Precision analysis of the proposed model across single and multi-word aspect terms.

- Yu et al. 2018 [76] proposed a model to apply a multi-task learning framework to implicitly capture the relations between aspect and opinion, by explicitly modeling syntactic constraints among aspect term extraction and opinion term extraction to uncover relationships. The proposers achieved an optimal solution over the neural predictions for both tasks. They claimed to be the first to explicitly model syntactic constraints in neural network-based approaches for aspect and opinion terms co-extraction.
- Agerri et al. 2019 [77] proposed a perceptron based algorithm and employed three groups of features—local shallow orthographic features, word shape and N-gram features, their context and simple clustering features based on unigram matching.
- Luo et al. 2019 [78] proposed a bidirectional dependency tree network that combines the information gained from bottom-up and top-down propagation on the given dependency syntactic tree. A complete framework is then developed to integrate the embedded representations and Bi-LSTM along with CRF to learn both tree-structured and sequential features to solve the aspect term extraction problem.
- Akhtar et al. 2020 [54] proposed a model which combines a Bi-LSTM and a CNN network for joint learning of aspect extraction and sentiment analysis. The Bi-LSTM module learns the sequential pattern of the sentence and predicts the boundaries of each aspect term in the sentence. The CNN module utilizes the local convolutional features of the aspect terms that are identified for sentiment classification. Both these modules work jointly and mutually maximize the prediction likelihood.

A comparison of the proposed model performance with the baselines considered is presented in Table 5. The best results reported on each dataset is marked in bold. The proposed model has reported the best F-measure of 0.81 on SemEval 2014 dataset among unsupervised models. It is almost on par with the recent supervised models too. With respect to SemEval 2015 dataset, the model has performed superior to the unsupervised baselines and has outperformed even the reported supervised models. The model is the second best on SemEval 2016 dataset with [Chauhan et al. 2020] reporting an F-measure of 0.79 on the dataset. Chauhan et al. 2020 actually employ a hybrid approach using a deep learning model which learns from the labels that

Table 5
Comparison of proposed model performance with supervised and unsupervised baselines.

Dataset	Methodology	SemEval 2014	SemEval 2015	SemEval 2016
Wu. Chuhan et al. 2018 [57]	Unsupervised	0.76	0.63	0.64
Dragoni et al. 2019 [73]	Unsupervised	–	0.60	0.67
Chauhan et al.2020 [58]	Unsupervised	–	–	**0.79**
Chen et al. 2017 [74]	Supervised	–	–	0.72
Wei et al. 2017 [75]	Supervised	0.83	0.67	0.72
Hu et al.2018 [53]	Supervised	–	–	0.74
Agerri et al. 2019 [77]	Supervised	0.84	0.71	0.74
Luo et al. 2019 [78]	Supervised	**0.85**	0.71	0.74
Yu et al. 2018 [76]	Supervised	0.84	0.71	–
Akhtar et al. 2020 [54]	Supervised	0.83	–	–
Proposed approach	Unsupervised	0.81	**0.74**	0.75

are predicted using a rule based approach in the first phase of their model and hence the comparative high performance. The proposed approach has outclassed all the reported unsupervised baselines across all datasets and the fact that it has stood on par with deep learning based supervised baselines is appreciable. [Luo et al. 2019] and [Aggeri et al. 2019] are a couple of supervised baselines which have reported a consistently high performance across all datasets. While the former incorporates the strength of dependencies into a deep learning model the latter is based on a simpler perceptron algorithm and has tried to capture the strength of semantics and contextual features.

The proposed enhanced guided LDA model has exhibited considerable improvement over the unsupervised baselines and is almost on par with the supervised baselines. The model has limitations and to understand the root cause a detailed error analysis is performed in the consequent subsection.

5.3. Error analysis

Investigating the uncovered aspect terms and the contributing factors to low precision is a good way of analyzing the model. The unreachable aspects could be either due to the guided restrictions imposed by the proposed model on the input sequences, the filters imposed by the model, the challenges imposed by the dataset or the dependency of the model on resources like POS taggers. Investigating in these directions, a categorized error analysis is presented.

5.3.1. Errors related to POS tags

A few among the unrecalled aspects lost owing to wrongly tagged or unreachable POS tags are listed in Table 6. They can be segregated into four categories.

Case 1: As observed in Table 6, the POS is wrongly tagged in the case of the aspect terms like '*tequila', 'omikase', 'hookah*' or *'naan'* and even multi-word aspects like *"king crab salad with passion fruit vinaigrette"*. All such cases of aspects are not reachable due to wrongly tagged POS tags are marked as *Case 1* in Table 6. This is a model drawback in terms of the resources it is dependent upon for generating the POS tags.

Case 2: In a few other cases presented in Table 6, like '*outside*' or *'billed'*, it is other POS tags marked as aspect terms which are very rare practically. A decision to consider terms corresponding to the regular expressions "VBN" or "RB" as prospective aspect terms cannot be incorporated in the model as it represents only a minority and would invite a large volume of false positives. Similarly, with respect to multi-word aspect terms "*Indo Chinese food*" and "*sitting space*" correspond to tag sequences "JJ* NN" or "VBG NN" which are uncovered by the proposed model as our model confines its filter only to highly reliable tags/tag sequences. Hyphenated aspects posed a challenge in SemEval 2015 datasets. The tag sequence "NN-NN" corresponding to the aspect term *"nagki-bokum"* in Table 6 could seem to be captured using the regular expression "N*" but hyphenations appear in the review sentences as separators for sentence clauses too, hence these hyphenated aspects are also uncovered by the model. All such cases of aspects not reachable due to uncovered POS tags are marked as *Case 2* in Table 6.

Case 3: A third category of unreachable tags include those corresponding to the regular expression "JJ NN*", but are filtered by the condition imposed by the proposed model that the token corresponding to *JJ* should not be an opinion word found in Bing Liu dictionary. A few examples in this category marked as *Case 3* in Table 6 is *"young woman", "hot dogs", "special roll"* etc. *Case 3* is the consequence of the restrictions imposed by the model for better precision.

Case 4: Often, aspects get engulfed in a larger POS pattern sequence and hence lost. For example, "*creme brulee*" corresponding to "NN NN" gets lost in a larger POS tag sequence "JJ NN NN" corresponding to "*silken creme brulee*". Similarly, "*mushroom sauce*" mapped to NN NN gets engulfed in *larger POS tag sequence* "NN IN NN NN" for the phrase "*maggot in mushroom sauce*". This category of error is mapped as *Case 4* in Table 6.

5.3.2. Errors related to aspects lost in semantic filter

A few among those aspects lost in semantic filters across all the datasets are listed in Table 7. These are lost for having a semantic similarity less than $\theta_{sem-low}$ (optimized value of 0.45) the lower semantic threshold, with the seed set. The usage of a context sensitive BERT model has been able to elevate the semantic similarities to a great extent. This is justified by an elevated semantic similarity of 0.351 for '*space*' and 0.385 for '*Bukhara*' which completely owes to context dependent distributional semantics which is unachievable by resources like WordNet or prior word embedding algorithms. The loss is minimal in terms of unrecalled aspects in the semantic filter when compared to the huge gain in filtering false positives.

5.3.3. Controversial labels in dataset

Controversial labels in the dataset also have been identified to a certain extent. A few examples to justify the argument are given in Table 8. For example, the term '*place*' is confusingly marked as a non-aspect term in spite of an opinion being expressed about the term '*place*' as noted in example 1 and 2. Similarly, in example numbered four in Table 8, '*chairs*' and '*table*' are marked as aspects in the dataset but the opinion holder is actually criticizing the bad service to which '*chairs*' and '*table*' are not even related implicitly. Similarly, in example 5 the complaint is about the '*waiter*' but the word '*kitchen*' *is* marked as an aspect term. In other examples in Table 8, "*parmesan cheese*", *'takeout', 'served'* etc. are marked as aspect labels but the corresponding examples do not convey any opinion on them.

Investigating the reasons for the dip in the precision of the proposed model, wrong tags leading to false positives is one reason which is minimal. The predominant cause is that the candidate aspects are semantically related to the seed words, but do not represent domain aspects or the review sentence does

Table 6
Unrecalled aspect terms (unidentified by POS tags).

Error category	SemEval 2014		SemEval 2015		SemEval 2016	
	Aspect term	POS tag	Aspect term	POS tag	Aspect term	POS tag
Case 1	Tequila	JJ	Hookah	JJ	Omikase	JJ
	Priced	VBN	Maître-D	NNP-FW	Naan	FW
	King crab salad with passion fruit vinaigrette	NN* IN NN* JJ			Chicken tikka masala	NN FW FW
Case 2	Middle eastern spreads	JJ JJ NN	Nagki-bokum	NN-NN	Eats	VBZ
	Iced blended mocha	JJ JJ NN	Four course prix fix menu	CD NN NN VB NN	Egg white omlette	NN JJ NN
	Outdoors	RB	Indo Chinese food	JJ* NN	House special roll	NN JJ NN
	Outside	RB	Dinner for two	NN IN CD	Filet mignon on top of spinach and mashed potatoes	JJ NN IN NN IN NN CC NN NN
	Billed	VBN	Selection of bottled beer	NN IN JJ NN	Bar keep	NN VBN
	Served	VBN	Sitting space	VBG NN	Waiting staff	VBG NN
Case 3	Hot dogs	JJ NN	Special roll	JJ NN	Young woman	JJ NN
	Artificial lobster meat	JJ NN*			Iced tea	JJ NN
Case 4	Creme brulee	NN*	Mushroom sauce	NN*	Seafood dinners	NN*

Table 7
Aspect terms unrecalled by the proposed model (lost in semantic filter).

SemEval 2014		SemEval 2015		SemEval 2016	
Aspect term	BERT based semantic similarity with seed words	Aspect term	BERT based semantic similarity with seed words	Aspect term	BERT based semantic similarity with seed words
Naan	0.342	Runner	0.32	Deck	0.409
Counter	0.341	Bark	0.235	Space	0.351
Port	0.238	Bukhara	0.385	Owner	0.215
Lighting	0.298	Venison	0.252	Artwork	0.239
Quesadilla	0.261	Boths	0.339	Apps	0.314
Attitude	0.401	Frontman	0.341	Standby	0.253
Texture	0.389	Mussles	0.403	Roe	0.357
Lounge	0.411	DJ	0.295	Toro	0.349
Rabbit	0.352	Kimchee	0.312	Bhartha	0.307
Dance floor	0.394	Patio	0.304	Open sesame	0.265
Jazz singer	0.310	Korean fair	0.354	Bus boy	0.219
Erbazzone emiliana	0.419	Village underground	0.231	Young woman	0.257

Table 8
A few controversial labels in SemEval datasets.

No	Sentence	Controversial label
1	I absolutely suggest this **place**	*place* is Non-aspect
2	If you re in New York you do not want to miss this **place**	*place* is Non-aspect
3	First walking in the **place** seemed to have great ambience	*place* is Aspect
4	then she made a fuss about not being able to add 1 or 2 **chairs** on either end of the **table** for additional people	*chairs* and *table* is Aspect
5	After waiting for almost an hour, the waiter brusquely told us he'd forgotten to give the **kitchen** our order	*kitchen* is Aspect
6	I went in one day asking for a **table** for a group and was greeted by a very rude hostess.	*table* is Aspect
7	Highly recommended... As stated, I haven't dined in the restaurant but stopped by there to pick up **takeout** and it seems a very relaxing place; also, the bar looks nice.	*takeout* is Aspect
8	They are **served** on Focacchia bread and are to die for.	*served* is Aspect
9	Another friend had to ask 3 times for **parmesan cheese**.	*parmesan cheese* is Aspect
10	last Tuesday for a **late lunch** with a friend	*late lunch* is Aspect

not contain an opinion towards the aspect term. An attempt has been made in the proposed model to incorporate this correlation of aspects and opinion words by incorporating opinion words (adjectives) in the model building. But multiple aspect terms in a review sentence and only a few mapped to opinion might be a challenging feature for LDA to learn. In certain cases, the false positive is a part of a multi-word aspect term partially extracted. For example, in the sentence *"The boutique selection of wines covers a wide variety without being imposeing"* the aspect term labeled is *"boutique selection of wines"* but the proposed model extracted only *"wines"*. Thus, partial recall of multi-word aspect terms is a major contributor to precision dip. To support this argument, an analysis of SemEval 2014 dataset reveals that *71%* of false positives in single word aspect terms are part of multi-word aspect terms. This also justifies the low recall for multi-word aspect terms. Considering that the failed recall in two word aspects is around 26% (as visible in Fig. 8), only 23% of those are completely unidentified, the remaining 77% are partially identified.

To conclude, wrong POS tags, lacking semantic similarity, partial recall in the case of multi-words aspect terms, lack of strength of the LDA model to a certain extent to differentiate non-opinionated aspect terms are the identified issues in the model. The major challenge lies with review sentences, which have multiple aspect terms and are all not being mapped to sentiment.

6. Conclusion and future work

The work proposes an enhanced guided LDA model for aspect term extraction in conjunction with BERT based semantic similarity. The model experiments the strength of guiding a topic model for aspect term extraction using minimal seeds which are representative of aspect categories. The input sequences are also filtered based on regular expressions and semantic and frequency filters. It is novel in that it has augmented semantic strength into the topic model in a unique way. The model performance has been compared with a few recent unsupervised and supervised baselines experimented on the same dataset and has reported appreciable results. The model has reported an F-measure of 0.81, 0.74 and 0.75 in the Restaurant domain of SemEval 2014, 2015 and 2016 datasets respectively. The proposed work has contributed an almost unsupervised model for aspect term extraction, which is on par with the state of art.

The scope for improvement is with respect to multi-word aspect terms. In the case of multi-word aspect terms, the unlabeled train data is not large enough to create a statistical significance for bi-grams or larger N-grams. If we could enlarge the train data corpora by appending unlabeled reviews from the same domain perhaps that would improve the correlation statistics of the aspect terms. There is a percentage of aspect terms that are lost in the semantic filter probably because their context in the review sentence could not create a domain representative vector representation. An improved method to choose a sentence that best explains the context could contribute to improve the semantic similarities. Deep Learning models are the state of art and have been exhibiting considerable performance on NLP tasks. A similar guided approach for a deep learning model which learns through distant supervision could be a future direction for exploration.

CRediT authorship contribution statement

Manju Venugopalan: Conceptualization, Methodology, Implementation, Visualization, Investigation, Data curation, Writing – original draft. **Deepa Gupta:** Conceptualization, Supervision, Validation, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] Bo Pang, Lillian Lee, Opinion mining and sentiment analysis, Found. Trends Inf. Retr. 2 (1–2) (2008) 1–135.

[2] Bing Liu, Sentiment analysis and opinion mining, Synth. Lect. Hum. Lang. Technol. 5 (1) (2012) 1–167.

[3] Housheng Xie, et al., A multi-dimensional relation model for dimensional sentiment analysis, Inform. Sci. 579 (2021) 832–844.

[4] Jesusa Serrano-Guerrero, Francisco P. Romero, Jose A. Olivas, Fuzzy logic applied to opinion mining: A review, Knowl.-Based Syst. (2021) 107018.

[5] Keith Cortis, Brian Davis, Over a decade of social opinion mining: a systematic review, Artif. Intell. Rev. 54 (7) (2021) 4873–4965.

[6] Meena Rambocas, Barney G. Pacheco, Online sentiment analysis in marketing research: a review, J. Res. Interact. Mark. (2018).

[7] Theodor Wyeld, et al., Understanding the effects of real-time sentiment analysis and morale visualisation in backchannel systems: a case study, Int. J. Hum.-Comput. Stud. 145 (2021) 102524.

[8] Ronen Feldman, Techniques and applications for sentiment analysis, Commun. ACM 56 (4) (2013) 82–89.

[9] Dongwen Zhang, et al., Chinese comments sentiment classification based on word2vec and SVMperf, Expert Syst. Appl. 42 (4) (2015) 1857–1863.

[10] B. Pang, L. Lee, S. Vaithyanathan, Thumbs up? Sentiment classification using machine learning techniques, EMNLP 10 (2002) 79–86.

[11] Alistair Kennedy, Diana Inkpen, Sentiment classification of movie reviews using contextual valence shifters, Comput. Intell. 22 (2) (2006) 110–125.

[12] Xiang Ma, et al., Rating prediction by exploring user's preference and sentiment, Multimedia Tools Appl. 77 (6) (2018) 6425–6444.

[13] Shi Feng, et al., Intersentiment: combining deep neural models on interaction and sentiment for review rating prediction, Int. J. Mach. Learn. Cybern. 12 (2) (2021) 477–488.

[14] Manju Venugopalan, G. Nalayini, G. Radhakrishnan, Deepa Gupta, Rating prediction model for reviews using a novel weighted textual feature method, in: Recent Findings in Intelligent Computing Techniques, Springer, Singapore, 2018, pp. 177–190.

[15] Swati Sanagar, Deepa Gupta, Automated genre-based multi-domain sentiment lexicon adaptation using unlabeled data, J. Intell. Fuzzy Systems 38 (2020) 6223–6334.

[16] Omid Mohamad Beigi, Mohammad H. Moattar, Automatic construction of domain-specific sentiment lexicon for unsupervised domain adaptation and sentiment classification, Knowl.-Based Syst. 213 (2021) 106423.

[17] Swati Sanagar, Deepa Gupta, Unsupervised genre-based multidomain sentiment lexicon learning using corpus-generated polarity seed words, IEEE Access 8 (2020) 118050-118071.

[18] Manju Venugopalan, Deepa Gupta, Exploring sentiment analysis on twitter data, in: Eighth International Conference on Contemporary Computing, IC3, IEEE, 2015, pp. 241–247.

[19] Manju Venugopalan, Deepa Gupta, Sentiment classification for hindi tweets in a constrained environment augmented using tweet specific features, in: International Conference on Mining Intelligence and Knowledge Exploration, Springer, 2015, pp. 664–670.

[20] Md Shahriare Satu, et al., Tclustvid: a novel machine learning classification model to investigate topics and sentiment in covid-19 tweets, Knowl.-Based Syst. 226 (2021) 107126.

[21] Anping Zhao, Yu Yu, Knowledge-enabled BERT for aspect-based sentiment analysis, Knowl.-Based Syst. (2021) 107220.

[22] Reinald Kim Amplayo, Min Song, An adaptable fine-grained sentiment analysis for summarization of multiple short online reviews, Data Knowl. Eng. 110 (2017) 54–67.

[23] Mohammad Ehsan Basiri, et al., A novel fusion-based deep learning model for sentiment analysis of COVID-19 tweets, Knowl.-Based Syst. 228 (2021) 107242.

[24] Mohammad Ehsan Basiri, Shahla Nemati, Moloud Abdar, Erik Cambria, U. Rajendra Acharya, ABCDM: An attention-based bidirectional CNN-RNN deep model for sentiment analysis, Future Gener. Comput. Syst. 115 (2021) 279–294.

[25] M. Usama, B. Ahmad, E. Song, M.S. Hossain, M. Alrashoud, G. Muhammad, Attention-based sentiment analysis using convolutional and recurrent neural network, Future Gener. Comput. Syst. 113 (2020) 571–578.

[26] Mohammed Arshad Ansari, Sharvari Govilkar, Sentiment analysis of mixed code for the transliterated Hindi and Marathi texts, Int. J. Nat. Lang. Comput. (IJNLC) 7 (2018).

[27] Rouzbeh Ghasemi, Seyed Arad Ashrafi Asli, Saeedeh Momtazi, Deep Persian sentiment analysis: Cross-lingual training for low-resource languages, J. Inf. Sci. (2020) 0165551520962781.

[28] Manisha Satish Divate, Sentiment analysis of Marathi news using LSTM, Int. J. Inf. Technol. 13 (5) (2021) 2069–2074.

[29] Navonil Majumder, Devamanyu Hazarika, Alexander Gelbukh, Erik Cambria, Soujanya Poria, Multimodal sentiment analysis using hierarchical fusion with context modeling, Knowl.-Based Syst. 161 (2018) 124–133.

[30] S. Poria, N. Majumder, D. Hazarika, E. Cambria, A. Gelbukh, A. Hussain, Multimodal sentiment analysis: Addressing key issues and setting up the baselines, IEEE Intell. Syst. 33 (6) (2018).
[31] E. Cambria, D. Das, S. Bandyopadhyay, A. Feraco, Affective computing and sentiment analysis, in: A Practical Guide to Sentiment Analysis, Springer, Cham, 2017, pp. 1–10.
[32] J. Han, Z. Zhang, N. Cummins, B. Schuller, Adversarial training in affective computing and sentiment analysis: Recent advances and perspectives, IEEE Comput. Intell. Mag. 14 (2) (2019) 68–81.
[33] E. Cambria, S. Poria, A. Hussain, B. Liu, Computational intelligence for affective computing and sentiment analysis [guest editorial], IEEE Comput. Intell. Mag. 14 (2) (2019) 16–17.
[34] Hu Minqing, Bing Liu, Mining and summarizing customer reviews, in: 10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2004, pp. 168–177.
[35] M. Venugopalan, D. Gupta, V. Bhatia, A supervised approach to aspect term extraction using minimal robust features for sentiment analysis, in: Advanced Computing and Intelligent Engineering, Springer, Singapore, 2021, pp. 237–251.
[36] S. Poria, E. Cambria, A. Gelbukh, Aspect extraction for opinion mining with a deep convolutional neural network, Knowl.-Based Syst. 108 (2016) 42–49.
[37] Manju Venugopalan, Deepa Gupta, An unsupervised hierarchical rule based model for aspect term extraction augmented with pruning strategies, Procedia Comput. Sci. 171 (2020) 22–31.
[38] Mohammad Tubishat, Norisma Idris, Mohammad Abushariah, Explicit aspects extraction in sentiment analysis using optimal rules combination, Future Gener. Comput. Syst. 114 (2021) 448–480.
[39] Jagadeesh Jagarlamudi, Hal Daume III, Raghavendra Udupa, Incorporating lexical priors into topic models, in: Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics, 2012, pp. 204–213.
[40] M. Hu, B. Liu, Mining opinion features in customer reviews, in: Proceedings of the 19th National Conference on Artificial Intelligence, AAAI, 2004, pp. 755–760.
[41] Z. Hai, K. Chang, J.-J. Kim, Implicit feature identification via co-occurrence association rule mining, in: International Conference on Intelligent Text Processing and Computational Linguistics, Vol. 6608, Springer, 2011, pp. 393–404.
[42] Z. Li, M. Zhang, S. Ma, B. Zhou, Y. Sun, Automatic extraction for product feature words from comments on the web, in: Asia Information Retrieval Symposium, Springer, 2009, pp. 112–123.
[43] Y. Zhao, B. Qin, S. Hu, T. Liu, Generalizing syntactic structures for product attribute candidate extraction, in: Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics, 2010, pp. 377–380.
[44] Toqir A. Rana, Yu-N. Cheah, A two-fold rule-based model for aspect extraction, Expert Syst. Appl. 89 (2017) 273–285.
[45] Q. Mei, X. Ling, M. Wondra, H. Su, C. Zhai, Topic sentiment mixture: modeling facets and opinions in weblogs, in: Proceedings of the 16th International Conference on World Wide Web, ACM, 2007, pp. 171–180.
[46] D.M. Blei, A.Y. Ng, M.I. Jordan, Latent dirichlet allocation, J. Mach. Learn. Res. 3 (2003) 993–1022.
[47] Ivan Titov, Ryan T. McDonald, A joint model of text and aspect ratings for sentiment summarization, ACL 8 (2008) 308–316.
[48] Ivan Titov, Ryan McDonald, Modeling online reviews with multi-grain topic models, in: Proceedings of the 17th International Conference on World Wide Web, ACM, 2008, pp. 111–120.
[49] Samuel Brody, Noemie Elhadad, An unsupervised aspect-sentiment model for online reviews, in: Annual Conference of the North American Chapter of the Association for Computational Linguistics, ACL, 2010, pp. 804–812.
[50] Ayoub Bagheri, Mohamad Saraee, Franciska De Jong, ADM-LDA: An aspect detection model based on topic modelling using the structure of review sentences, J. Inf. Sci. 40 (5) (2014) 621–636.
[51] H. Ye, Z. Yan, Z. Luo, W. Chao, Dependency-tree based convolutional neural networks for aspect term extraction, in: Pacific-Asia Conference on Knowledge Discovery and Data Mining, Springer, 2017, pp. 350–362.
[52] Lei Shu, Hu Xu, Bing Liu, Controlled cnn-based sequence labeling for aspect extraction, 2019, arXiv preprint arXiv:1905.06407.
[53] Xu Hu, Bing Liu, Lei Shu, Philip S. Yu, Double embeddings and cnn-based sequence labeling for aspect extraction, ACL 2 (2018) 592–598.
[54] Md Shad Akhtar, Tarun Garg, Asif Ekbal, Multi-task learning for aspect term extraction and aspect sentiment classification, J. Neurocomput. 398 (2020) 247–256.
[55] Ling Luo, Xiang Ao, Yan Song, Jinyao Li, Xiaopeng Yang, Qing He, Dong Yu, Unsupervised neural aspect extraction with sememes, IJCAI (2019) 5123–5129.
[56] Yukun Ma, Haiyun Peng, Tahir Khan, Erik Cambria, Amir Hussain, Sentic LSTM: a hybrid network for targeted aspect-based sentiment analysis, Cogn. Comput. 10 (4) (2018) 639–650.
[57] C. 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.
[58] G.S. Chauhan, Y.K. Meena, D. Gopalani, R. Nahta, A two-step hybrid unsupervised model with attention mechanism for aspect extraction, Expert Syst. Appl. 161 (2020) 113673.
[59] A. Giannakopoulos, C. Musat, A. Hossmann, M. Baeriswyl, Unsupervised aspect term extraction with b-lstm & crf using automatically labeled datasets, in: Proceedings of the 8th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis, ACL, 2017, pp. 180–188.
[60] Zhaoxia Wang, Seng-Beng Ho, Erik Cambria, Multi-level fine-scaled sentiment sensing with ambivalence handling, Int. J. Uncertain. Fuzziness Knowl.-Based Syst. 28 (04) (2020) 683–697.
[61] Zhaoxia Wang, et al., Anomaly detection through enhanced sentiment analysis on social media data, in: 2014 IEEE 6th International Conference on Cloud Computing Technology and Science, IEEE, 2014, pp. 917–922.
[62] Erik Cambria, et al., SenticNet 6: Ensemble application of symbolic and subsymbolic AI for sentiment analysis, in: Proceedings of the 29th ACM International Conference on Information & Knowledge Management, 2020, pp. 105–114.
[63] Wei Li, Luyao Zhu, Erik Cambria, Taylor's theorem: A new perspective for neural tensor networks, Knowl.-Based Syst. 228 (2021) 107258.
[64] Wei Li, et al., BiERU: bidirectional emotional recurrent unit for conversational sentiment analysis, Neurocomputing 467 (2020) 73–82.
[65] Ziyu Zhou, Qianqian Wang, R-transformer network based on position and self-attention mechanism for aspect-level sentiment classification, IEEE Access 7 (2019) 127754-127764.
[66] Qian Liu, et al., Automated rule selection for aspect extraction in opinion mining, in: Twenty-Fourth International Joint Conference on Artificial Intelligence, 2015, pp. 1291–1297.
[67] Devlin Jacob, Ming-Wei Chang, Kenton Lee, Kristina Toutanova, Bert: Pre-training of deep bidirectional transformers for language understanding, in: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 2018, pp. 4171–4186.
[68] Garima Dwivedi, Manju Venugopalan, Deepa Gupta, A statistical-semantic PSO model for customer reviews-based question answering systems, in: International Conference on Soft Computing and Signal Processing, Springer, Singapore, 2019.
[69] Á. Elekes, A. Englhardt, M. Schäler, K. Bohm, Toward meaningful notions of similarity in NLP embedding model, Int. J. Digit. Libr. 21 (2) (2020) 109–128.
[70] Pontiki Maria, Dimitrios Galanis, John Pavlopoulos, Harris Papageorgiou, Ion Androutsopoulos, Suresh Manandhar, Semeval-2014 task 4: Aspect based sentiment analysis, in: Proceedings of the International Workshop on Semantic Evaluation, SemEval 2014, 2014, pp. 27–35.
[71] Pontiki Maria, Dimitrios Galanis, Haris Papageogiou, Suresh Manandhar, Ion Androutsopoulos, Semeval-2015 task 12: Aspect based sentiment analysi, in: Proceedings of the International Workshop on Semantic Evaluation, SemEval 2015, 2015, pp. 486–495.
[72] 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, Semeval-2016 task 5: Aspect based sentiment analysis, in: 10th International Workshop on Semantic Evaluation, 2016, pp. 19–30.
[73] Mauro Dragoni, Marco Federici, Andi Rexha, An unsupervised aspect extraction strategy for monitoring real-time reviews stream, Inf. Process. Manage. 56 (3) (2019) 1103–1118.
[74] 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.
[75] Wei Xue, W. Zhou, T. Li, Q. Wang, MTNA: a neural multi-task model for aspect category classification and aspect term extraction on restaurant reviews, in: Proceedings of the Eighth International Joint Conference on Natural Language Processing, 2017, pp. 151–156.
[76] Jianfei Yu, Jing Jiang, Rui Xia, Global inference for aspect and opinion terms co-extraction based on multi-task neural networks, IEEE/ACM Trans. Audio, Speech, Lang. Process. 27 (1) (2018) 168–177.
[77] Rodrigo Agerri, German Rigau, Language independent sequence labeling for opinion target extraction, Artificial Intelligence 268 (2019) 85–95.
[78] H. Luo, T. Li, B. Liu, B. Wang, H. Unger, Improving aspect term extraction with bidirectional dependency tree representation, IEEE/ACM Trans. Audio, Speech, Lang. Process. 27 (7) (2019) 1201–1212.