Knowledge and Information Systems (2022) 64:1625–1651
https://doi.org/10.1007/s10115-022-01678-5

REGULAR PAPER

Unsupervised model for aspect categorization and implicit aspect extraction

Omar Mustafa AL-Janabi[1] · **Nurul Hashimah Ahamed Hassain Malim**[1] · **Yu-N Cheah**[1]

Received: 26 May 2021 / Revised: 23 February 2022 / Accepted: 26 February 2022 /
Published online: 25 April 2022

Abstract

People's ability to quickly convey their thoughts, or opinions, on various services or items has improved as Web 2.0 has evolved. This is to look at the public perceptions expressed in the reviews. Aspect-based sentiment analysis (ABSA) deemed to receive a set of texts (e.g., product reviews or online reviews) and identify the opinion-target (aspect) within each review. Contemporary aspect-based sentiment analysis systems, like the aspect categorization, rely predominantly on lexicon-based, or manually labelled seeds that is being incorporated into the topic models. And using either handcrafted rules or pre-labelled clues for performing implicit aspect detection. These constraints are restricted to a particular domain or language which is domain-dependent. In this work, we first propose a novel unsupervised probabilistic model Topic-seeds Latent Dirichlet Allocation (TSLDA) that leverages semantic regularities for the articulation of explicit aspect-categories. Then, based on the articulated categories, a distributed vector is used for the identification of implicit aspects. The experimental results show that our approach outperforms baseline methods for different domain-data with minimal configurations. Specifically, utilizing the RI measure, our proposed TSLDA outperformed multiple clustering and topic models by an average of 0.83% in diverse domain-data, and roughly 0.89% using the Precision metric for implicit aspect detection.

Keywords Topic-seeds · Topic modelling · Word embedding · Aspect categorization · Sampling algorithm · Implicit aspects

1 Introduction

The emergence of Social Web serves as a platform where thousands of consumers are able to publish their product reviews online. The contents of these reviews are usually not explicitly machine-processable. Sentiment analysis (SA) or opinion mining is a type of automated analysis that recognizes and extracts affective states and subjective information from text. It uses computational methods and natural language processing to identify and extract content indi-

✉ Nurul Hashimah Ahamed Hassain Malim
nurulhashimah@usm.my

1 School of Computer Sciences, Universiti Sains, 11800 George Town, Pulau Pinang, Malaysia

rectly and delivers human consumable information [1, 2]. Aspect-based sentiment analysis is a vital subtask in sentiment analysis due to its paramount importance in extracting the aspects (attributes or components of products of the reviews) from the reviews and determining the sentiments (usually positive, negative or neutral) reflected towards it. The aspect-based analysis is a fine-grained task that breaks down the text to phrase-level to be more concise about one's satisfaction with the product or service. The phrase-level in aspect-based sentiment analysis can be classified into explicit aspects that analyze aspect terms explicitly mentioned within the sentence, and implicit aspects which are embedded (have hidden meaning) in the context of the sentence.

Many previous studies have analyzed explicit aspects, while a limited few tackled the implicit aspects. Researchers were not interested in implicit aspects because it requires natural language understanding rather than natural language processing [3]. In brief, it relies on the clue of the context (because the implicit aspects are hidden within the context), while explicit aspects are more word-based. For instance, the word "affordable" in "the camera is affordable" is an implicit aspect referring to the aspect "price" of the camera. Based on previous studies, several methods have been proposed to tackle the explicit aspect categorization and the implicit aspect identification. A notable method for aspect categorization is the probabilistic method that relies on the lexicons and manually labeled seeds in guiding the distribution of the words [4–6]. The contribution of this study is twofold. The first contribution is the Topic-seeds Latent Dirichlet Allocation (TSLDA) model to extract and categorize explicit aspects, in which the semantic regularities model is proposed to maintain coherent topics distribution in Latent Dirichlet Allocation (LDA) [7]. The second is the adoption of distributed vector (Skip-gram) for the extraction of the implicit aspects. Technically, Skip-gram uses NCE algorithm as computational optimization method to update the output vector of the model per iteration.

TSLDA differs from other related work in that, we are proposing a domain-trained word embedding approach which is sufficient to guide the model without the need for manually labeled seeds, or those lexicon-based methods such as WordNet that requires a regular update for newly generated terms. This makes TSLDA better than other related works which could be time inefficient and domain-dependent because they need to be generated according to the domain dataset. For the implicit aspect detection, methods such as handcrafted rules, co-occurrence, association rules, clustering and distance knowledge have been used. In terms of time efficiency and domain dependencies, the distance knowledge from the web is the most convenient to be used where the similarity between terms is measured based on their hits on the internet as in [4]. However, the hits values vary according to the retrieving time [8]. Therefore, we propose a distributed vector (DV), known as the Skip-gram model for implicit aspect detection. Skip-gram is an unsupervised neural network model that is capable of determining word similarities without jeopardizing the semantic of the words. This eliminates the reliance on web knowledge (internet hits) for implicit aspect detection. This manuscript is organized as follows; the first section covers the introduction, the second section brings forth research objectives, the third section is recent work, the fourth section presents the proposed method, the fifth section is evaluation of the contribution and finally the conclusion.

2 Research objectives

The exponential growth of online reviews requires automated text processing techniques to be developed for transforming the words into insights, that would be useful for the public and merchants. The best method would be the one that could suit different domain (domain-independent). In aspect-based categorization, approaches that are cost-effective and considered semantic regularities are required to accomplish several tasks related to analyzing and summarizing the opinionated reviews. A plethora of supervised machine learning has given promising achievements in aspect-level analysis for aspects extraction, categorization, and implicit aspects extraction. Yet supervised methods require a lot of labeled data and that is time inefficient. In addition, several probabilistic models advise manually labeled seeds and lexicons for the categorization and classification of the aspect-terms in sentiment analysis, and that leads to the problem of domain-dependency as these lexicons (like WordNet) have to be updated frequently with newly generated terminologies. Similarly, handcrafted rules that are used in related works for the identification of the implicit aspects require new rules for the new domains and languages in text data. To these issues our objectives in this work are:

- To propose an unsupervised model that can be used for aspect-based tasks without the need for labeled data for training.
- To propose domain-trained word embedding to consider the semantic regularities between words for aspect categorization.
- To propose Skip-gram as a distributed vector for the identification of the implicit aspects.

3 Recent work

In data mining, the problem of assigning text reviews into categories is known as text classification (or text categorization) [9]. Whilst, in the field of sentiment analysis, aspect categorization is the grouping of the aspect-terms into categories mostly accomplished using topic models (aka probabilistic topic models[10–15]).

3.1 Aspect categorization methods

Probabilistic topic models have been presented as a unification model for extracting and categorizing aspects collectively [15, 16]. Technically, the typical LDA model suffers from the lack of semantic regularities between topics and is ineffective for phrase-based/aspect-based analysis because it was designed for long documents and huge corpora. Therefore, lexicon-based approaches [4] and the word segmentation procedure [17] are being added to LDA. Using lexicon-based methods to create topic seeds leads to domain-dependent issues, in which each new domain or set of data requires its own set of related seeds.

Ontology-based semantic knowledge is introduced to enhance the performance of topic models. Santosh et al. [18] proposed feature ontology tree (FOT) on the LDA topic modeling that specifies topics by randomly assigning the word's vocabulary of documents to the topics by using Dirichlet distribution. Thereafter, the topics to the words are assigned using the term frequency-inverse document frequency (TF–IDF) value in the form of word probability. Ontology-based inspired, Ali et al. 2019 [19] first, generated latent topics using standard LDA, then manually collected transportation knowledge from transport-related websites using an ontology-based query. Then, they filtered each LDA-based generated topic using an ontology-based similarity measure.

Other researchers considered the word embedding models to improve the LDA-based generated topics. Park et al. [20] proposed LDA to generate latent topics of main aspects for a particular product, then clean irrelevant aspects using word embedding and clustering algorithm. Likewise, an unsupervised model was developed [21] to deal with different domains and languages. They proposed the bootstrapping model that relies on word embedding, brown clustering, and MaxEnt to classify the sentiment and the aspect words.

Notwithstanding, because the LDA method fails to perform well in phrase-based /aspect-based analysis. By aggregating short texts to construct faux lengthy documents, the co-occurrence pattern of a term in the document is presented to solve the data sparsity problem. In their work on Twitter analysis, Nimala et al. [22] applied this method by combining hashtag-based tweet aggregation strategy. For short text reviews, Xiong et al. introduced the WSTM (Word-pair Sentiment-Topic Model) [23]. They used WSTM to represent the generation process of sentiment and aspect terms at the same time. They employed a sliding window in which they created word pairs at each step and then incorporated these word pairs in their topic model. Tang et al. introduced the JABST (joint aspect-based sentiment topic) model, which simultaneously models aspects, opinions, sentiment polarity, and granularities to extract multi-grained aspects and views [24]. However, Ozyurt and Akcayol [17] propose a frequency-based method for extracting frequent nouns as aspects and association rules for segmenting sentences, as well as confidence methods for attaching segmented nouns to their related-words and mapping segments into aspect categories based on semantic similarity.

3.2 Implicit aspect extraction methods

The extraction of implicit aspects can be categorized into: Unsupervised, Supervised, and Semi-supervised, as can be found in Tubishat, Idris, & Abushariah [25].

Gobi and Rathinavelu [26] introduced the co-occurrence method for implicit aspects extraction that relies on the co-occurrence association between the opinion words in the implicit reviews and the extracted explicit aspects which are noun phrases.

Twofold co-occurrence methods [27] proposed for the extraction of implicit aspects using co-occurrence matrix rely on the ontology class of the sentiment-words. First fold, sentiment words that are adjectives, adverbs, and verbs are extracted, then using the co-occurrence matrix, the highest occurrence of aspect with these sentiment words, is extracted as the aspect-term. The extraction relies on the explicit aspects that are extracted following the ontological basis between the opinion-target and the sentiment-words. Hence, a co-occurrence matrix is developed to include the explicit aspects and their related sentiment-words, while the implicit aspects are indicated using the sentiment-words in the reviews that have no explicit aspects related to them. Rana and Cheah [28] extracted explicit aspects and used them to extract the implicit aspects by building a co-occurrence matrix between the explicit aspects and the sentiment-words. The sentiment-words in the implicit reviews were assigned into the explicit aspects in the co-occurrence matrix based on the semantic similarity between the sentiment-words and the explicit aspects. A proper similarity score was taken using the Normalized Google Distance (NGD) algorithm. Feng et al. [29] performed the extraction of implicit aspects using the co-occurrence matrix between the sentiment-words and the explicit aspects. For the sentiment-words in the reviews that were not related to any explicit aspects, the words are being compared with the annotated explicit aspects. The highest occurrence between the sentiment-word and the explicit aspects is considered as implicit aspect.

Afzaal et al. [30] proposed the decision tree algorithm for the extraction of the implicit aspects that are in unigram segmentation form. The extraction first relies on the extracted explicit aspects using the rule-based method. Then the sentiment-words in the implicit reviews are assigned into the related classes of explicit aspects using decision tree algorithm.

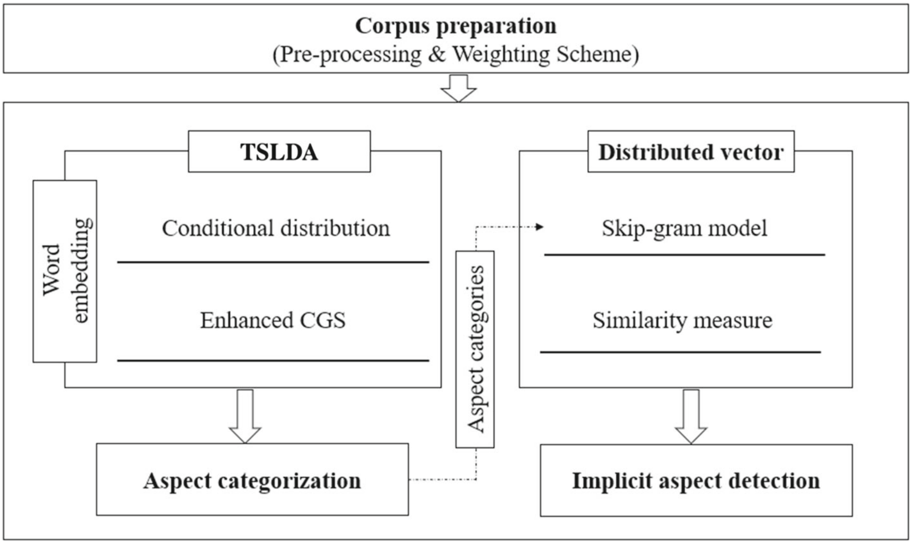

Fig. 1 The proposed framework for aspect-based sentiment analysis

Mowlaei et al. [31] proposed a genetic algorithm with a lexicon-based method for the extraction of the implicit aspects. Particle swarm optimization (PSO) was used as a classifier on the MPQA, SentiWordNet, and Bing Liu's lexicons for implicit extraction. The non-negative matrix factorization (NMF) method was proposed in [32] for the extraction of the implicit aspects. This work differs from the previous work in that, we attained the semantic regularities between the words in conditional distribution and the sampling algorithm of the topic model, and we attained the implicit aspects extraction with no rule-based methods.

4 Proposed framework

This section shows the proposed framework to achieve the articulated objectives (Sect. 2). The proposed framework in Fig. 1 advises unsupervised probabilistic model (TSLDA), and distributed vector (Skip-gram) for the explicit aspect categorization, and implicit aspects identification, respectively.

4.1 Probabilistic topic model

LDA is a well-known probabilistic model that relies on the Bag of Words (BOW) to perform the distribution of words to topics. Relying on the BOWs for aspect categorization ignores the low-frequent words, yet, dismisses the semantic association of the words. Other methods have been previously proposed to cope with the issue by formulating handcrafted rules and seeds [25]. It is time-inefficient domain-dependent because the crafted rule has to be recrafted for the new domains and languages, as well as for the manually labeled seeds. Therefore, we have improved the performance of LDA by incorporating external knowledge into the distribution of the words that led to a new model named TSLDA model. The external pieces of knowledge are prior knowledge from the word embedding model, that is generated using the domain-trained datasets. The next section explains the topic seeds retrieval in word embedding before getting into topic modeling.

4.1.1 Topic seeds retrieval

To avoid computational difficulty such as incoherent topic modelling, most of the recent efforts have empowered the distribution of the words to their topics by exploiting manual-labeling of seeds or dictionaries. This measure has led to the issue of domain-dependent. Therefore, we have introduced the TSLDA model to incorporate prior-knowledge using word embedding (WE). That is the Continuous Bag of Words (CBoWs)[1] model trained using our domain-data (e.g. SemEval-2014 dataset). The prior knowledge includes the topic-seeds that had been retrieved using the WE. The most similar words can be retrieved using similarity measure using word embedding. Cosine similarity (cosine) was used to find the similarity between two vectors. It is a similarity measure between two vectors a, and b of an inner product space that measures the cosine of the angle between them. The most similar words can be retrieved by taking the max cosine similarity as follows: the dot product is calculated as

$$a\ .\ b = \sum_{i=1}^{N} a_i b_i \tag{1}$$

the norm is defined as

$$||x|| = \sqrt{x.x}, \tag{2}$$

while the cosine similarity measure is defined as

$$\text{cosine}(a, b) = \frac{a.b}{||a|| * ||b||} \tag{3}$$

which given Eqs. (1) and (2) becomes

$$\text{cosine}(a, b) = \frac{\sum_{i=1}^{N} a_i b_i}{\sqrt{\sum_{i=1}^{N} a_i^2}\ \sqrt{\sum_{i=1}^{N} b_i^2}} \tag{4}$$

Formula (4) applied to a pair of N-dimensional vectors, that is two vectors a, and b in our case.

This part of the work retrieved the topic seeds based on the number of topics in the dataset, for instance, four topic seeds were retrieved for the "SemEval 2014 Restaurant". Each topic-seed was used to guide the topic distribution of the aspect-terms. In other words, the semantically related aspect-terms was assigned to the aspect-category that was closely related in terms of semantic association. In our model TSLDA, the aspect categories (or topics) represent the aspect terms that are semantically related to each other, i.e. the aspect term "price", "cheap", "overpriced" and "cost" are in the aspect category that is "price". Therefore, the sampling algorithm collapsed Gibbs sampling (CGS) had been enhanced by adapting prior knowledge from word embedding.

[1] CBoWs is a trainable word embedding model chosen to maintain the semantic of the words in LDA and to retrieve the topic seeds based on the semantic similarity measure

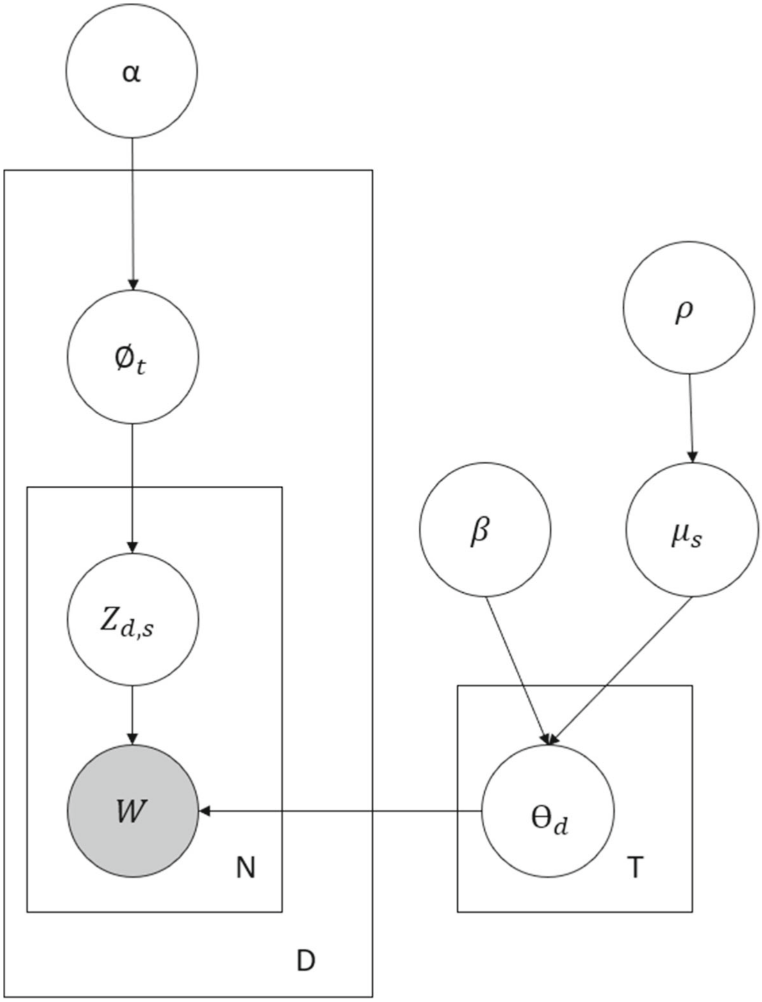

Fig. 2 The graphical model of TSLDA

4.1.2 TSLDA generative process

Figure 2 gives the graphical representation of our proposed TSLDA model that is described as follows: for each document d ∈ {1, ...D}, a distribution of topics, θ_d, is being sampled from a Dirichlet distribution (Dir) that is represented by the parameter α_d, which is presented as a vector that governs the distribution of topic priors for the documents. Each document is referred to as a single review sentence in the corpus, because we are dealing with a fine-grained analysis of aspects. While, each sentence $s_d \in \{1, \ldots, s_d\}$, has a single aspect, and a distribution of topics, ϕ_t, also sampled from Dirichlet distribution (Dir) that is represented by parameter β, which governs the distribution of aspect-terms to the aspect-category. For each aspect-term (word) $w \in \{w_1, \ldots, w_w\}$ in a sentence s_d, a topic distribution ϕ_t, is biased by topic-seeds, μ_s. These seeds are sampled using hyperparameter ρ, which is a vector that tunes the distribution of word to topic or aspect-category.

To discover the aspect categories, topic to document distribution is performed using multinomial distribution. While word to topic distribution is intractable, we have introduced Collapsed Gibbs sampling (CGS) algorithm that is enhanced by embedding topic-seeds into the conditional distribution to tune parameter β, which governs word to topic distribution. The enhanced CGS gives a complete description of the conditional assignment probability of a single-word topic assignment, conditioned to the rest of the model. Given the assignment of latent variables, topic assignment $Z_{d,n}$, of each word in the document are semantically assigned to the related aspect-category. The following equation represents the conditional distribution of enhanced word to topic distribution.

$$p\left((z_{d,n} = t\ \right) - z_{d,n}\ ,\ w; \alpha,\ \beta,\ \rho\ ) \propto \frac{(n_t^{(d)} + \alpha)}{(n_{\cdot}^{(d)} +\ Ta)} * \frac{(n_t^{(w)} + \beta)}{(n_t^{(.)} +\ w\beta)} * \frac{(n_s^{(d)} +\ \rho)}{(n_t^{(.)} +\ s\rho)} \tag{5}$$

where, $n_t^{(d)}$ is the number of times an aspect-term from document d that was assigned to aspect-category, $n_t^{(w)}$ is the number of times aspect-term was assigned to aspect-category, and $n_t^{(w)}$ is the number of times a seed from topic-seeds in document d is assigned to aspect-category.

4.1.3 An improved collapsed Gibbs sampling

The improved CGS has leveraged the semantic similarity measure to draw the word to topic cognitively. In each d $\in \{1\ldots D\}$ document, every $w_{d,n}$ in the currently assigned $z_{d,n}$ aspect-category, within i number of iterations, decreasing the variable associated with the topic/aspect-category assignment. The modified CGS assures the coherency of the topics based on the given threshold of similarity between the current word and topic-seeds that have been initially incorporated. If we considered the lowest threshold for similarity is 0.2, unlikely the current $w_{d,n}$ will be sampled to the $z_{d,\ n}$, thus, β is set to a small value that prevents the sampling of the words to the aspects category (e.g. the small β value is 0.1). Otherwise, if the semantic similarity of the current $w_{d,n}$ is $\geq$ 0.2, it is most likely to be sampled to the $z_{d,\ n}$, if so, β value will be augmented to attempt the sampling (e.g. the value of β augmented up to 0.7), to assign the current word to its aspect-category, and increment of the topic assignment again.

4.2 Implicit aspect detection

Figure 3 illustrates the overall procedure for the detection of implicit aspect. The proposed model is divided into three main steps: 1) the generated aspect categories using TSLDA being fed into the DV (skip-gram model), 2) query the opinionated implicit aspect reviews, and 3) retrieve the implicit aspects to their vector space.

4.2.1 Distributed vector

The procedure of training for the proposed Skip-gram model starts by applying one-hot encoding on the training data, as the text data cannot be sent directly to the matrix. That means we have a vector of all the unique words in the training data, as the length of the vector equals the total number of tokens (the sentence reach the maximum of 20 tokens). The

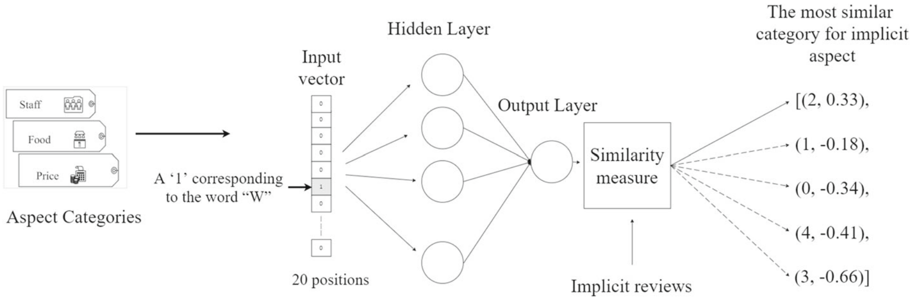

Fig. 3 The proposed model for implicit aspect detection

encoding of each token corresponds to one in the vector and zero everywhere. Since we are implementing the skip-gram model, the encoding of the words is represented with window size. For instance, for the target [natural] and context (language, processing), the encoding is [1, 0, 0, 0, 0, 0, 0], for the target and [[0, 1, 0, 0, 0, 0, 0], [0, 0, 1, 0, 0, 0, 0]] for the context, as the sentence starts with "natural language processing" etc.

The procedure of the one-code is applied in the aspect-categories to return the one-code encoding of each aspect.

Using the training data of aspect categories obtained from TSLDA, we have trained the skip-gram word embedding model. As it is a neural network, two weight matrices being initialized: one is the context matrix and the other is the embedding matrix.

The hidden layer in the trained model is generated by performing a dot product between the context matrix and the target word w_t.

Skip-gram uses the w_t *t*o predict the surrounding words "w_{t+j}". The objective of the Skip-gram, however, is to sum the log probabilities of the words on the left and the right of the target/centre word w_t, to produce the following objectives:

$$J_\theta = \frac{1}{T}\sum_{t=1}^{T}\sum\nolimits_{-n\le j\le n,\ne 0} \log p\left(w_{t+j}\middle|\ w_t\right) \tag{6}$$

The computation of the $P\left(w_{t+j}\middle| w_t\right)$ performed using SoftMax function:

$$p\left(w_t\middle| w_{t-1},\ \ldots,\ w_{t-n+1}\right) = \frac{\exp\left(h^T v'_{w_t}\right)}{\sum_{w_i\in V}\exp\left(h^T\ v'_{w_i}\right)} \tag{7}$$

Rather than computing the probability of the target word w_t given its previous words, the probability of a surrounding word w_{t+j} given w_t has been calculated. The variables in the above equation have been replaced with:

$$p\left(w_{t+j}\middle| w_t\right) = \frac{exp(h^T v'_{w_{t+j}})}{\sum_{w_i\in V}\ exp(h^T\ v'_{w_i})} \tag{8}$$

In the SG model, h parameter is the word embedding v_{w_t} of the input word w_t , besides, the centre word replaced with w_I and the surrounding words with w_O:

$$p\left(w_O\middle| w_I\right) = \frac{exp(v_{w_t}^T v'_{w_{t+j}})}{\sum_{w_i\in V}\ exp(v_{w_t}^T\ v'_{w_i})} \tag{9}$$

Nevertheless, to force the output of the neural network to sum to one, several methods have been implemented. For instance, a square error measure for training a neural net has been used. However, the drawbacks of the square error measure are depriving the network if probabilities are being assigned to the mutually exclusive class label. Further, the SoftMax function which is a soft continuous version of the maximum function has been used as a hierarchical SoftMax function in [33, 34]. It is a binary Huffman tree, where short words are assigned to frequent words. But in our work, we have proposed Noise Contrastive Estimation (NCE) [15]. The underlying idea of NCE is that the functional model should be able to differentiate data from noise by the leverage of logistic regression. Besides, it can also be used to maximize the log probability of the SoftMax function approximately. For the proposed model, there are two vector representations for each word within the vocabulary: the input vector V_w and the output vector $V_w^{'}$. Learning the input vector is cheap, but it is expensive for the output vector. This is where we have used the NCE algorithm as computational optimization so that many output vectors could be updated per iteration. This is despite the fact that, the hierarchical SoftMax tries to estimate the probability of the word directly. Instead, NCE uses an auxiliary loss that optimizes the probability of correct words. For every word w_i given it is context c_i of n previous words $w_{t-1}, \ldots, w_{t-n+1}$ in the training set. K noise samples $\widehat{w}_{ik}$ thus generated from Q. Unigram distribution has been employed for a generation, and labels for the correct and noise words have been used to differentiate the correct and the corrupt class in the algorithm calculated using equation 10. In the class label, the correct words are designated with $y = 1$ and $y = 0$ for others.

$$\widehat{w}_{ik} \sim Q \tag{10}$$

Logistic regression is used to minimize the negative log-likelihood, i.e. cross-entropy of our training data against the noise:

$$J_\theta = -\sum_{w_i \in V} [\log P\,(y = 1|\, w_i,\ c_i) + kE_{\widehat{w}_{ik}\sim Q}\left(\log P(y = 0|\widehat{w}_{ij},\ c_i)\right)] \tag{11}$$

4.2.2 Query implicit review

Implied aspect terms are not explicitly mentioned in the reviews. It can be detected if the reviews satisfy two conditions:

- The aspect-terms does not appear in the review explicitly
- It can be implied by opinion-words or the surrounding words of the sentence.

The opinion-words are usually expressed using Adjectives or Adverbs part of speech. For instance, the implicit aspect can be inferred from sentence (1) using the indicator word "half off" which is referring to the aspect category "price".

<sentence id="1">The drinks are amazing and half off till 8pm.
<sentence id="2">Also they were $15 each!.

Yet, the inferring of the implicit aspects sometimes requires the whole review to infer the implicit aspect term, as in sentence (2) where the whole review referring to the aspect category price.

As shown in the example above, there is no clue in the review referring to the aspect category price except the meaning of the whole sentence. To this end, we have proposed the distributed vectors that accept variable length to detect the implicit aspect reviews in online reviews.

For every sentence in the corpus (e.g. SemEval 2014: Restaurant), there are three conditions that needed to be true for an opinion word to be considered as an implicit aspect. These conditions are: (i) existence of opinion word within the sentence; (ii) the sentence review had no explicit aspect-terms assigned to them and; (iii) the sentence was assigned into an aspect-category. For instance, 'the food is affordable', suppose the review sentence was assigned into an aspect-category 'price', and there was no explicit aspect-terms assigned to them but there was an opinion-word existed ('affordable'). The sentence would be considered an implicit aspect review that had to be categorized into the suitable aspect-category, which was the aspect-category 'price' in this case.

In other scenarios, there were more than one opinion words within the sentence review, and these opinion words were referring to two different aspect-categories. Thus the algorithm would check the opinion words that were being assigned into aspect-terms, if the opinion words had no aspect-terms, then the opinion words would be assigned into the semantically related aspect-category separately.

However, there were implied reviews that had the preposition 'it' in them, and they had not been assigned into aspect-terms, yet they were assigned into the aspect-category. Such kinds of implied reviews were regarded as a complete sentence to be located into their related aspect-category.

4.2.3 Vector space of implicit review

The retrieval of semantically related aspect-category relies on the vector space of the acquired implicit review. The query relies on the distributed representation and the semantic similarity measures that are presented in [35]. Cosine Similarity is a similarity measure used to find the similarity between two $N_{dimension}$ vectors a and b. The similarity is calculated as in equation 4.

The detection of implicit aspects relies on the semantic measure trained using the content of the aspect-categories. Hence, the relatedness to the aspect-category is retrieved by exploiting the semantic similarity between the implicit reviews and the trained document vector. Here, the implicit review was regarded as a query using infer-vector against the aspect-category that had aspect-terms within them. For instance, the implicit aspect sentences 'This isn't a value joint.', and 'Also they were $15 each!' formed the restaurant domain and were assigned into the aspect-category 'price', as illustrated in Table 1-the cosine measure used to determine the degree of similarity between the query and the aspect-category. The semantically similar aspect-category to the implicit aspects is ranged from 'zero' (dissimilar) to 'one' (highly similar). The section below shows the effectiveness of the proposed model.

5 Evaluation

5.1 Dataset

Manually annotated datasets were chosen to evaluate the performance of the proposed framework. The first dataset TripAdvisor's Hotel[2] is annotated into seven aspect categories (i.e. "Room", "Value", "Location", "Cleanliness", "Check_in", "Service", "Business"). The other chosen datasets come from International Workshop on Semantic Evaluation (SemEval). SemEval datasets focus on Aspect-based sentiment analysis (ABSA) tasks. In SemEval-2014[3] Restaurant dataset, aspect term extraction conducted to identify the aspect terms

[2] https://www.cs.cmu.edu/~jiweil/html/hotel-review.html.

[3] https://alt.qcri.org/semeval2014/task4/.

Table 1 Categorizing implicit aspects

Review sentence	Aspect-category
"This isn't a value joint"	Price
"Also they were $15 each!"	Price

Table 2 Datasets characteristics

Dataset	Short form	#Sentences	# Aspect-category
TripAdvisor: Hotel	"TA:H"	587095	7
SemEval-2014 Restaurant	"R:14"	3041	5
SemEval-2015 Restaurant	"R:15"	2048	5
SemEval-2016 Restaurant	"R:16"	2288	6

and further the aspect terms are annotated into five distinct aspect categories (i.e. "food", "service", "price", "ambience", "anecdotes/miscellaneous"), the expressed reviews in the dataset revolving around the restaurant's qualities in terms of the annotated aspect categories. SemEval-2015[4] is a continuation of the SemEval-2014, which has reviews on the Restaurant domain that have their target aspect annotated into categories. Restaurant dataset contains features (e.g. "restaurant", "food", "drinks", "service", "ambience", and "location") annotated into 5 attribute labels (i.e. "general", "prices", "quality", "style_options", and "miscellaneous"). The Restaurant reviews come from SemEval-2016[5] annotated into 6 attributes (i.e. "Service", "Restaurant", "Food", "Location", "Ambience", "Drinks"). Table 2 shows the number of sentence reviews and categories in each dataset.

5.2 Evaluation measure

Rand Index (RI), Entropy, Normalized Mutual Information (NMI), and Purity as evaluation metrics were used to evaluate the performance of the presented models for the detection of the aspect categories (aspect categorization). RI [36], Entropy [37], NMI [38], and Purity [36] measured the similarity between the aspect-category "ground truth label" $G_{Gold} = g_1, g_2, g_3, \ldots, g_n$ from the dataset and the generated aspect-category $D_{model} = d_1, d_2, d_3, \ldots, d_n$ of our model. For instance, in TripAdvisor: Hotel dataset, there were seven ground truth aspect-category: Rooms, Cleanliness, Value, Service, Location, Check-in, and Business.

The gold standard aspect-categories were annotated in the original dataset, in which there were 20 aspect-terms in each category. And the number of aspect-category has relied on the used dataset.

On the other hand, the identification performance of implicit aspects evaluated using precision, recall, and f-score. Precision: is the fraction of the relevant implicit aspect among the retrieved implicit aspects. Recall: is the fraction of the total amount of the relevant implicit aspects that were actually retrieved. While F-score is a measure of test accuracy.

4 https://alt.qcri.org/semeval2015/task12/.

5 https://alt.qcri.org/semeval2016/task5/.

5.3 Experiments and results

The experimental design of our framework dealt with a series of experiments with different settings of the proposed TSLDA, and Skip-gram models.

TSLDA model encompasses three separate methods that have different hyperparameters that must be tuned:

- **Meth 1**: Hyperparameters of the unsupervised word embedding models (Dt-WE).
- **Meth 2**: Hyperparameters of the term weighting scheme (feature extraction) methods.
- **Meth 3**: Hyperparameters of the Bayesian model (TSLDA).

Meth1 (i.e. Dt-WE) had been leveraged to be incorporated into TSLDA, to semantically guide the distribution of the words in the model. By incorporating the topic seeds into the conditional distribution of the model. Further, the incorporation of Meth 1 enhanced the performance of the sampling method using similarity measure.

Meth 2 (e.g. count vectorizer (CV), and Term Frequency-inverse document frequency (TF-IDF)) as a feature extraction method used to represent the document-words as a sparse matrix.

In the proposed word embedding model, 'Meth 1' comprises three main hyperparameters; window size, dimensions, and epochs, that can have sets of values for each hyperparameter. Following [39], window size being first initialized with '5', while it is '300' for the number of dimensions. However, in Meth 2, the hyperparameters of (CV, TF-IDF) that must be tuned are the Max-df, Min-df, and Max-features. Max-df is the maximum required occurrence of the vocab in the documents, Min-df is the minimum required occurrence of the vocab in the documents, and Max-features is the maximum number of unique words. While for the other hyperparameters in Meth 2 (i.e. CV, and TF-IDF) like the stop-words, lowercase, token-pattern, has been set to English, True, and [a-zA-Z0-9]{3,} respectively. The range of the values of the TSLDA hyperparameters was determined based on the recently proposed topic models [31-34].

Based on the values of the stated hyperparameters for Meth 2 and Meth 3, a Grid-Search method was conducted to find out the values of the optimal hyperparameters. The Grid-Search methods selected the best values based on the best Log-Likelihood score of TSLDA.

Tables 3, 4, 5, and 6 reports the obtained results from applying different configurations for the categorizations of the aspects in several datasets (including, SemEval-2014 Restaurant '**R:14**', SemEval-2015 Restaurant '**R:15**', SemEval-2016 Restaurant '**R:16**' TripAdvisor: Hotel '**TA:H**'). The applied configurations are:

(1) Multi-Topic Seeds (M-TS + GS): meaning that, we have exploited a set of topic seeds[6] for each category for the guidance of the Conditional Distribution. These topic seeds being retrieved using the Word Embeddings model. To do so, we used two different Word Embeddings models i) Domain-trained word embeddings (Dt-WE). It was trained using domain-data (e.g. R:14), the domain-data being pre-processed before the training of Dt-WE. ii) GloVe word embeddings as a pre-trained word embedding was also utilized to be compared with Dt-WE while they were being incorporated into TSLDA. Further, in 'M-TS + GS' configuration, we used Gibbs Sampling (GS) algorithm for the sampling of the words in the TSLDA model. GS is a standard sampling algorithm that has no word embedding.

[6] Note: Topic-seeds: is a set of terms/seeds for each category (e.g. 'price', 'money', and 'pay' are the topic-seeds for the aspect category 'price').

Table 3 Compare the **RI** score of different configurations for TSLDA in terms of topic seeds, sampling algorithm, feature extraction for aspects categorization

Configuration	WE	FE	R:14	R:15	R:16	TA:H
M-TS + GS	Dt-WE	CV	0.5907	0.5702	0.5953	0.5400
		TF-IDF	0.2358	0.4623	0.2430	0.3589
	GloVe	CV	0.5978	0.5784	0.5953	0.5178
		TF-IDF	0.3799	0.4956	0.2430	0.4765
S-TS + GS	Dt-WE	CV	0.5789	0.4252	0.5708	0.5487
		TF-IDF	0.3684	0.3955	0.2134	0.3718
M-TS + CGS	Dt-WE	CV	0.7812	**0.8308**	**0.8198**	0.8705
		TF-IDF	0.6166	0.6766	0.6732	0.8579
	GloVe	CV	0.7536	0.7887	0.7859	0.8558
		TF-IDF	0.7536	0.7807	0.7856	0.8378
S-TS + CGS	Dt-WE	CV	**0.8025**	0.8150	0.7932	**0.8789**
		TF-IDF	0.7801	0.7760	0.7939	0.8175
	GloVe	CV	0.7906	0.7880	0.7102	0.8412
		TF-IDF	0.7536	0.6977	0.6945	0.7635

Bold values indicate the best results

Table 4 Compare the **NMI** score of different configurations for TSLDA in terms of topic seeds, sampling algorithm, feature extraction for aspects categorization

Configuration	WE	FE	R:14	R:15	R:16	TA:H
M-TS + GS	Dt-WE	CV	0.253	0.278	0.292	0.251
		TF-IDF	0.021	0.123	0.126	0.235
	GloVe	CV	0.253	0.225	0.274	0.302
		TF-IDF	0.021	0.125	0.012	0.087
S-TS + GS	Dt-WE	CV	0.124	0.125	0.125	0.215
		TF-IDF	0.102	0.098	0.070	0.123
M-TS + CGS	Dt-WE	CV	**0.421**	**0.482**	**0.486**	**0.656**
		TF-IDF	0.232	0.241	0.244	0.654
	GloVe	CV	0.284	0.335	0.236	0.554
		TF-IDF	0.284	0.331	0.333	0.412
S-TS + CGS	Dt-WE	CV	0.401	0.442	0.401	0.601
		TF-IDF	0.335	0.301	0.386	0.412
	GloVe	CV	0.335	0.301	0.333	0.478
		TF-IDF	0.294	0.252	0.201	0.332

Bold values indicate the best results

(2) Single-Topic Seeds (S-TS + GS): we used a Single seed/term for each aspect category (e.g. topic seed 'price' is for the aspect category 'price' in the R:14). In this configuration, S-TS[7] investigated with standard GS.

(3) Multi-Topic Seeds accompanied by Collapsed Gibbs Sampling (M-TS + CGS): meaning that, M-TS being exploited and CGS was used as a sampling algorithm which was the

[7] Note: S-TS: is a short form for Single Topic-seed: meaning single seed or term for each category.

Table 5 Compare the **Entropy** score of different configurations for TSLDA in terms of topic seeds, sampling algorithm, feature extraction for aspects categorization

Configuration	WE	FE	R:14	R:15	R:16	TA:H
M-TS + GS	Dt-WE	CV	1.6122	1.7025	1.7652	1.7254
		TF-IDF	2.1458	1.8325	2.2148	2.1002
	GloVe	CV	1.6154	1.7854	1.8531	2.0024
		TF-IDF	2.1254	2.1042	2.1251	2.1201
S-TS + GS	Dt-WE	CV	1.7685	1.8423	1.7235	1.2025
		TF-IDF	2.2152	2.3457	2.3824	2.2823
M-TS + CGS	Dt-WE	CV	1.1325	**0.9762**	**1.0001**	0.9504
		TF-IDF	1.4685	1.1215	1.2214	0.9985
	GloVe	CV	1.2546	1.1252	1.2214	0.9754
		TF-IDF	1.4872	1.4212	1.4253	0.9975
S-TS + CGS	Dt-WE	CV	**1.0875**	1.0215	1.0985	**0.9572**
		TF-IDF	1.1212	1.1215	1.1211	1.0251
	GloVe	CV	1.2300	1.1254	1.2356	09962
		TF-IDF	1.2354	1.6423	1.2380	1.2543

Bold values indicate the best results

Table 6 Compare the **Purity** score of different configurations for TSLDA in terms of topic seeds, sampling algorithm, feature extraction for aspects categorization

Configuration	WE	FE	R:14	R:15	R:16	TA:H
M-TS + GS	Dt-WE	CV	0.631	0.610	0.616	0.666
		TF-IDF	0.407	0.452	0.432	0.235
	GloVe	CV	0.633	0.423	0.432	0.485
		TF-IDF	0.220	0.385	0.232	0.235
S-TS + GS	Dt-WE	CV	0.663	0.392	0.423	0.452
		TF-IDF	0.365	0.333	0.341	0.353
M-TS + CGS	Dt-WE	CV	0.623	**0.688**	**0.587**	0.701
		TF-IDF	0.622	0.651	0.544	0.694
	GloVe	CV	0.602	0.610	0.574	0.631
		TF-IDF	0.612	0.593	0.582	0.600
S-TS + CGS	Dt-WE	CV	**0.670**	0.684	0.531	**0.712**
		TF-IDF	0.614	0.596	0.444	0.620
	GloVe	CV	0.668	0.631	0.580	0.696
		TF-IDF	0.672	0.644	0.454	0.675

Bold values indicate the best results

improved sampling algorithm. In which we incorporated two different Word Embeddings model for two separate implementations on each dataset. That was Dt-WE incorporated into the CGS, and results of the aspects categorization obtained for each dataset. On the other hand, the GloVe model was incorporated into the TSLDA model to perform the categorization. This is to compare the performance of TSLDA while using 'Dt-WE' and while using a pre-trained word embedding 'GloVe'. Note that the optimal hyperparameters were selected for the Dt-WE based on the outcome of Dt-WE on TSLDA for each dataset, and the outcome of that was assessed using Log-Likelihood.

(4) Single-Topic Seeds accompanied by Collapsed Gibbs Sampling (S-TS + CGS): in this configuration, a single topic seed for each category was incorporated into the conditional distribution of the TSLDA model. While the Collapsed Gibbs sampling had been used as a sampling algorithm for the proposed models 'TSLDA'. However, 'S-TS' configuration had been used with different term weighting schemes (i.e. CV, and TF-IDF) and different word embedding (I.e. Dt-WE, and GloVe) while the sampling algorithm used was CGS.

Multiple -Single Topic-seed for each aspect-category of each dataset using Domain-trained word embeddings (Dt-WE) and GloVe pre-trained embeddings.

Further, the best **RI, NMI, Entropy, and Purity** score for the coverage of the aspect-terms in each dataset is highlighted in **bold font**. Out of the presented configurations for the categorization of the aspects, the results obtained using Dt-WE are relatively better than the use of pre-trained model 'GloVe' while conducting the fully developed model TSLDA as in Tables 3, 4, 5 and 6. The evaluation score for the TSLDA reported according to the selected configuration: Word Embeddings, feature extractions (weighting scheme), and the developed portions of the TSLDA. The performance of using multiple topic seeds with the GS for both the Dt-WE and GloVe produced high accuracy for the CV weighting scheme than TF-IDF. Regardless of that, the performance of M-TS + GS using Dt-WE achieved comparatively better accuracy than using GloVe. On the other hand, the use of Single topic seeds accompanied by GS being used with Dt-WE only, as the outcomes of using different word embeddings obtain the same results as using Dt-WE for such configuration 'S-TS + GS'. The fully developed configurations, which was the use of the enhanced conditional distribution, and the enhanced CGS, produced the highest accuracy using original components of the topic modelling like conditional distribution and GS algorithm. Hence, in terms of NMI score, the performance of using M-TS + CGS using Dt-WE and CV methods achieved the highest accuracy than the use of GloVe and the other FE/weighting scheme for all datasets as in Table 4. In terms of RI, Purity, and Entropy, Tables 4, 6 and 7 state the performance of using M-TS + CGS using Dt-WE and CV methods achieved the best accuracy for R:15, and R:16 datasets. While, it was best accuracy in terms of RI, Purity, and Entropy for the R:14 and TA:H datasets when using S-TS + CGS configuration as in Tables 3, 5 and 6. Nonetheless, the performance of the proposed TSLDA model based on the modified component(s) is depicted in Figs. 4, 5, 6 and 7. On the R:15 and R:16 datasets, Figs. 4, 6 and 7 demonstrate how the RI, Entropy, and Purity scores steadily increased while using the M-TS + CGS configuration, peaking at the employment of DT-WE and CV. For the TA:H and R:14 datasets, however, it is better to use the S-TS + CGS setup. Similarly, for the implemented datasets as shown in Fig. 5, the NMI score is the best when employing the M-TS + CGS setup.

5.4 Compared models

The compared models are the traditional clustering algorithms (i.e. Kmeans, NMF, and LSI) and the state-of-the-art topic models. Comparing to the previous methods, TSLDA outperformed most of the current methods in terms of aspects categorization:

Kmeans [40]: a clustering method that is commonly used to automatically generate the number of groups based on the distributional similarity.

O-LDA [13]: is an improved topic model that stands for the online variational inference as a sampling algorithm and Latent Dirichlet allocation topic model.

G-LDA [14]: is a topic model that used the Gaussian distribution to maintain the semantic regularities.

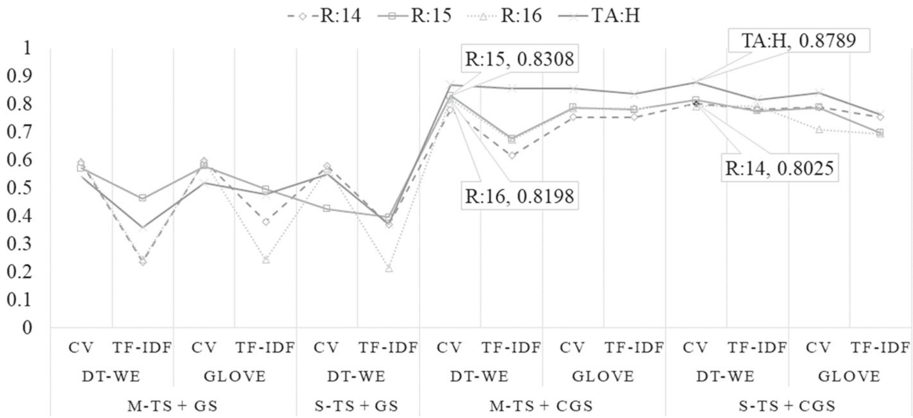

Fig. 4 Compare the **RI** score of different configuration for **TSLDA** Model

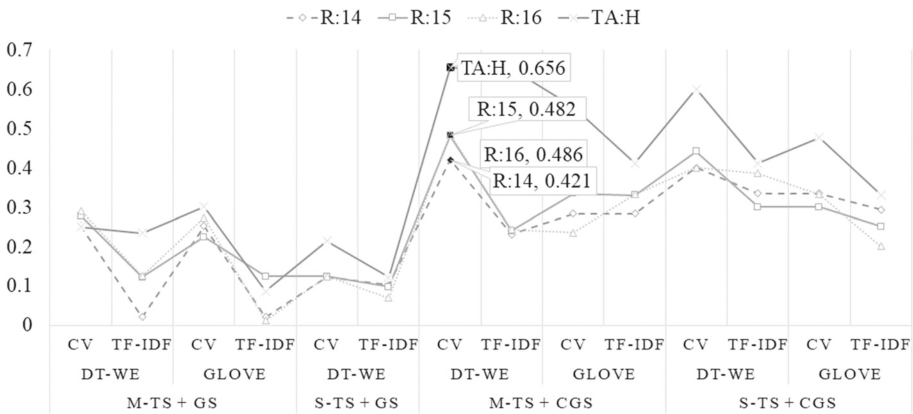

Fig. 5 Compare the **NMI** score of different configuration for **TSLDA** Model

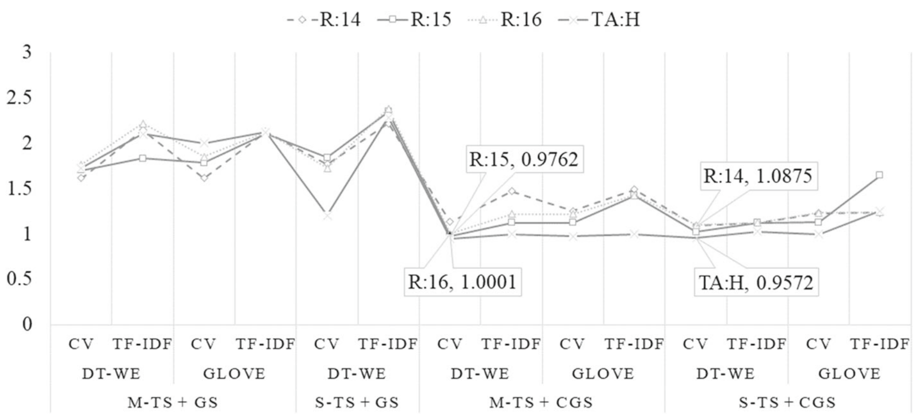

Fig. 6 Compare the **Entropy** score of different configuration for **TSLDA** Model

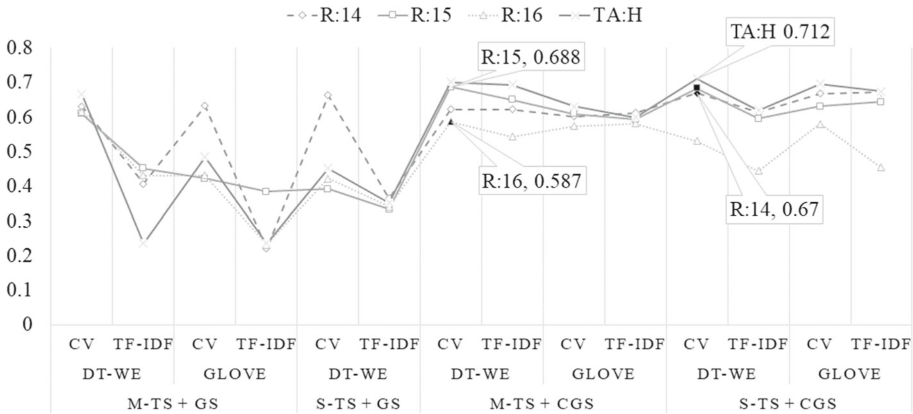

Fig. 7 Compare the **Purity** score of different configuration for **TSLDA** Model

Table 7 Extracted aspects in R:14 dataset using TSLDA model

Topic 0	Topic 1	Topic 2	Topic 3
Customer service	Lobster sandwich	Deal	Ambience
Waitress	Flavor	**Simply**	atmosphere
Reservation	Indian food	Reasonable	Neighborhood
Check	Fried chicken	**Friendly**	Music
Arrive	Mushroom	**List**	**Authentic**
Staff	Portion size	Overall	Décor

NMF [41]: is a clustering algorithm used to find a group of topics based on the linear representations of non-negative data.

LSI [42]: Latent Semantic Indexing is used to generate the group of topics.

LDA [7]: is the standard LDA model that is built based on the conditional distribution and the Gibbs sampling method.

The coverage of the extracted aspects is a factor to evaluate the trained model. Table 7 stated the top extracted aspect terms in R:14 dataset. The **bolded** aspect terms are the mistakenly extracted aspects in each category. The evaluation methods (RI, NMI, Entropy, and Purity) assess the performance of the trained models by calculating the correctly extracted aspects compared to the mismatched aspects.

Tables 8, 9, 10 and 11 show that TSLDA outperformed clustering algorithms and topic models that have an improved sampling algorithm. The performance of the current methods for aspect categorization is relatively affected by the aspect ‘price’, the synonyms aspects for this aspect are poorly retrieved using implemented topic models and clustering algorithms. For instance, in the restaurant domain dataset “R:14”, the aspect terms for the aspect category “price” are expressing the price of the food, which are difficult for the topic models to group their synonyms aspects into the same category.

As compared to baseline methods, on both the clustering algorithms and topic models, TSLDA shows promising performance over the baselines on average (avg%[8]) of 0.83%,

[8] Note: avg% is the sum of the averaged percentage values of the utilised evaluation techniques (e.g., 0.83 in the last row in Table 8 is the performance of TSLDA’s RI score on all four data sets).

Table 8 Comparison of **RI** with baselines

	RI				
	R:14	R:15	R:16	TA:H	Avg%
Kmeans	0.643	0.593	0.526	0.67	0.60
O-LDA	0.610	0.610	0.562	0.63	0.60
G-LDA	0.780	0.760	0.744	0.81	0.77
NMF	0.744	0.625	0.682	0.78	0.71
LSI	0.595	0.550	0.623	0.67	0.61
LDA	0.550	0.372	0.390	0.571	0.47
TSLDA	**0.80**	**0.83**	**0.81**	**0.87**	**0.83**

The best results are in bold

Table 9 Comparison of **NMI** with baselines

	NMI				
	R:14	R:15	R:16	TA:H	Avg%
Kmeans	0.093	0.066	0.037	0.208	0.10
O-LDA	0.095	0.215	0.012	0.227	0.14
G-LDA	0.278	0.421	0.256	0.402	0.34
NMF	0.3	0.111	0.304	0.409	0.28
LSI	0.028	0.3	0.1	0.341	0.19
LDA	0.121	0.121	0.115	0.144	0.12
TSLDA	**0.421**	**0.482**	**0.486**	**0.656**	**0.51**

The best results are in bold

0.51%, 1.00%, and 0.67% using RI, NMI, Entropy, and Purity metrics on all the trained datasets respectively. Besides, the NMF method has overcome the compared clustering algorithms such as Kmeans, and LSI. Because NMF relies on the linear representations of non-negative data to group the aspects into categories. Whilst, G-LDA as an improved topic model has given better results than standard LDA and improved O-LDA topic model. Because of that, the G-LDA model is adapting the semantic regularities between the topics and generated the aspect categories based on the usage of Gaussian distribution. Notwithstanding, both the proposed model TSLDA and G-LDA utilize word embeddings to maintain the semantic regularities between the topics, yet the TSLDA model has produced finer modelling of the aspects because the latter is using topic-seeds retrieval technique to guide the distribution of the aspect to the relevant topic(s).

Tables 12, 13 and 14 compare aspect categorization performance using TSLDA model with the baselines on the R:15, R:16 and TA:H datasets, respectively. The compared baselines for aspect grouping in R:15 dataset are the following:

MRA [43]: mutual reinforcement approach.

CAFÉ [44]: applies agglomerative clustering to construct aspect categories based on the word frequency.

NN-M [45]: Non-negative matrix factorization for aspect clustering.

In Table 12, CAFÉ as a clustering algorithm achieved better results than those reported in NN-M, and MRA, this is due to two reasons: first, frequency-based method is used to select a candidate aspects based on the word frequency, second, a domain-dependent similarity measure is proposed to identify the aspects that belongs to the same cluster. Further, CAFÉ

Table 10 Comparison of **Entropy** with baselines

	Entropy				
	R:14	R:15	R:16	TA:H	Avg%
Kmeans	1.32	2.013	1.741	1.021	1.52
O-LDA	1.332	1.452	1.4	1.006	1.30
G-LDA	1.21	1.001	1.181	0.989	1.10
NMF	1.32	1.136	1.119	0.99	1.14
LSI	2.125	1.458	1.359	1.032	1.49
LDA	1.768	1.784	1.725	1.2	1.62
TSLDA	**1.087**	**0.976**	**1.000**	**0.957**	**1.00**

The best results are in bold

Table 11 Comparison of **Purity** with baselines

	Purity				
	R:14	R:15	R:16	TA:H	Avg%
Kmeans	0.467	0.489	0.432	0.511	0.47
O-LDA	0.45	0.43	0.436	0.421	0.43
G-LDA	0.595	0.591	0.594	0.532	0.58
NMF	0.625	0.522	0.442	0.622	0.55
LSI	0.478	0.364	0.511	0.421	0.44
LDA	0.669	0.492	0.472	0.604	0.56
TSLDA	**0.679**	**0.688**	**0.587**	**0.712**	**0.67**

The best results are in bold

Table 12 Comparison with baselines on R:15 dataset

	MRA	CAFÉ	NN-M	TSLDA
RI	0.61	0.68	0.62	0.83

also overcome the MRA, even though, MRA used opinion words in the sentences to identify the aspect candidate that latter helped in clustering of the aspects (aspect categorization). Using no aspect candidate nor frequency-based methods, TSLDA outperform the advised methods by 0.22%, 0.15%, and 0.21% on MRA, CAFÉ and NN-M correspondingly.

Aspect categorization baselines on the R:16 dataset are the following:

FC-Kmeans [46]: this is also another semantic relevance-based constrained K-Means algorithm with weighted meaning representation.

FFDML [47]: this a neural network method implemented by fusing concepts from a knowledge base with the context information for the task of aspect grouping. **AP** [48]: this method only uses aspect term embedding without context information as feature vectors for K-means.

L-EM [48]: this is an unsupervised clustering algorithm that employed lexical knowledge for aspect grouping.

The presented neural network model "FFDML" in Table 13 comparatively gave better results compared to the rest of the methods (i.e., FC-Kmeans, AP, and L-EM) that relied on the kmeans clustering algorithm for aspect categorization. As FFDML benefited from using the context information of the sentences and interpolated them as an attention mechanism layer into the constructed neural network, the attention mechanism helped in find the semantic

Table 13 Comparison with baselines on R:16 dataset

	RI	NMI	Entropy	Purity
FC-Kmeans	0.5802	0.0544	1.6592	0.617582
AP	0.5676	0.0541	1.6783	0.6064
FFDML	0.5983	0.0568	1.6107	**0.6305**
L-EM	0.5537	0.0452	1.7048	0.6099
TSLDA	**0.81**	**0.486**	**1.000**	0.587

similarities between the aspects that are grouped in to categories. Yet TSLDA overcome the performance of FFDML in 21%, 0.43%, and 0.61% for the RI, NMI and entropy, except for that, in terms of purity metric, FFDML is higher than TSLDA by 0.05%.

However, the following are the compared topic models with TSLDA on the TA:H dataset:

DF-LDA [12]: Dirichlet Forest-LDA is a semi-supervised model which allows the distribution based on pre-stated constraint for each topic.

SAS [11]: Seeded aspect and Sentiment model: is a topic model which relies on seed sets for the distribution of the topics.

SenticLDA [4]: an unsupervised model which relies on a lexicon-based named (SenticNet).

SJASM [15]: Supervised joint aspect and sentiment model: represents each review document in the form of opinion pairs.

sLDA [49]: Supervised LDA: weakly supervised joint sentiment-topic model.

JST [16]: joint sentiment-topic model: a sentiment layer being combined into LDA.

ASUM [50]: aspect and sentiment unification model: leveraged sentiment lexicon.

LARA [51]: Latent aspect rating analysis: identify the semantic aspects based on the regression model.

Table 14 shows a comparison of state-of-the-art topic models with the proposed TSLDA model. SAS, DF-LDA, and SenticLDA were chosen as the best state-of-the-art aspect categorization topic models. The modelling of DF-LDA is limited by domain constraints, such as Must-link and Cannot-link constraints, which effectively limit the distribution of aspects in the relevant topics/categories. For the newly trained domain-data, new constraints are required. The seeded aspects and sentiment words were also used by SAS to manage the topic distribution. SenticLDA, on the other hand, is considered to use a lexicon-based approach and manually labelled seeds to maintain semantic regulations and control topic distribution. On the TA:H dataset, SenticLDA outperformed the SAS and DF-LDA models, as well as the baselines sLDA, JST, ASUM, and LARA. The given model TSLDA outperformed all of the other models and produced comparable results to SenticLDA. Aside from that, certain modern approaches, such as [4, 11, 12], were created to deal with aspect words with single segmentation (Uni-gram). Their models were used to extract aspects with a single segmentation (for example, aspect terms in the 'TA:H' dataset), whereas our model was used to conduct aspect terms in different segmentation (for example, Uni-gram, Bi-gram, and Tri-gram) using a single topic seed for each category as shown in Table 7.

The produced aspect categories using TSLDA were integrated into the distributed vector for the implicit aspect identification. The proposed distributed vector instantiated with a vector size of 20 words and the model was iterated over the aspect-categories over 50 times.

The adaptation of NCE to approximately maximize the log probability of the SoftMax function had improved the performance of the exploited distributed vector representation.

Table 14 Comparison with baselines on TA:H dataset

	DF-LDA	SAS	SenticLDA	sLDA	JST	ASUM	LARA	TSLDA
RI	0.75	0.77	0.86	0.72	0.70	0.69	0.63	**0.87**
Entropy	1.75	1.45	0.95	–	–	–	–	**0.95**

Table 15 Comparison of implicit aspect extraction method on the R:14 dataset

	Precision	Recall	F-score
CNN	0.8827	0.861	0.8815
CRF	0.8535	0.8272	0.8401
Rule-based	0.8521	**0.8815**	**0.8665**
Ours (Skip-gram)	**0.8914**	0.7808	0.8324

The best results is in bold

The detection of the implicit aspects has been conducted on two datasets (i.e. R:14, and R:15), while the following are state-of-the-art references for implicit aspects:

CNN [52]: Classification algorithm.
CRF [39]: Classification algorithm.
Rule-based [53]: Handcrafted rules.
CoAR [54]: Co-occurrence matrix and association rule.
CRSA [55]: Co-occurrence frequency.
CBA [56]: Classification method.
NCBA [57]: Co-occurrence matrix.
EXPRS [57]: Dependency relations.
CWC [52]: Context weight co-occurrence matrix.

As can be seen from Table 15, our proposed model (Skip-gram) produced comparative accuracy in terms of Precision results in comparison to the state-of-the-art methods for implicit aspect identification on R:14 dataset. Poria et al. [39] stated that the proposed supervised neural network CNN achieved 0.88 and 0.86 percent for precision and recall, respectively. While Toh & Wang[58] proposed a supervised machine learning using conditional random field (CRF) to extract the aspect terms. CRF produced a precision of 0.85 %, while recall was 0.82 %. However, Poria et al.[53] proposed the handcrafted rules (Rule-based) to extract the aspects based on the clues and linguistic patterns of the formulated aspects in a review. The performance of the extraction was 0.85 % for the precision and 0.88 % for the recall making it the highest recall.

Table 16 demonstrates that our method has the highest precision among others. Our method outperforms the co-occurrence matrix-based approaches CoAR, NCBA, CWC, and N-NM by a wide margin. Our approach, for example, has a precision of 0.89 %, which is higher than CoAR, NCBA, CWC, and N-NM, which have precision of 0.75 %, 0.74 %, 0.74 %, and 0.76 %, respectively. The reason for this is that our method makes use of a lot of information from the review text. The implicit aspect mapping to the relevant aspect category is performed using the similarity measure, and the distributed vector is trained using the generated aspect categories from TSLDA. Furthermore, we can see that CWC has the lowest recall. The reason behind this is because CWC considers all corpus nouns with a lot of noisy data to be candidate aspects. EXPRS only slightly outperforms CWC because it extracts explicit aspect words and opinion words based on only four dependency links, ignoring other relations in reviews. CoAR also fails because it only constructs association rules between opinion words

Table 16 Comparison of implicit aspect extraction method on the R:15 dataset

	Precision	Recall	F-score
CoAR	0.7571	0.4741	0.5831
CRSA	0.7689	0.9854	**0.8638**
CBA	0.7605	**0.9919**	0.8610
NCBA	0.7443	0.6828	0.7122
EXPRS	0.7331	0.4623	0.5670
CWC	0.7407	0.2913	0.4181
N-NM	0.7661	0.9887	0.8633
Ours (Skip-gram)	**0.8914**	0.7808	0.8324

The best results is in bold

and aspect words if their frequencies exceed the support and confidence thresholds, hence excluding some accurate but rare rules. While the highest recall goes to CBA as it grouped characteristics that appeared in the same category as the same opinion word. Then, in the test stage, they created a classifier to map a sentence into one aspect-opinion pair. Besides, as seen in Table 16, our method does not do well in recall. Our technique, for example, came in fourth place, after CBA, CRSA, and N-NM. The three techniques produced higher recall because they recognise sentences as implicit sentences with a high degree of reliability. CBA, for example, only considers sentences with two express opinion words, resulting in a high recall rate. However, there are no specific requirements for words in sentences in our approach. One cause could be that when the review set is tiny, some aspect words and opinion words are uncommon. There may be some implied terms that are overlooked, resulting in a tiny number of correctly identified sentences.

This research proposed an unsupervised model that can be used to analyze people's opinions on a product quality or services. In the sustainable Development goals,[9] the proposed model can be used to analyze the poverty-related opinions[10] from the web. Therefore, these reviews can be analyzed to study the poverty situation in regard to the cause of poverty or hanger. A study conducted by [59], they have studied the importance of poverty in the sustainable development goals. Other studies are used to conduct an online reviews in regard to the corruption and violence in the third world countries.[11,12] The October 2021 elections in Iraq are shrouded in uncertainty, as armed political groups[13] pressurise Iraqis to vote for a specific party under duress or bribe, resulting in a corrupted country that will always stay corrupted under these conditions. The proposed model in this work can be used to analyze an Iraqi collected online opinions regarding the transparency of the held election.

6 Conclusion

In this work, we present a novel framework which is based on the developed TSLDA and distributed vector (Skip-gram) model for aspect categorization and implicit aspect detection tasks. Experiments using the most recent available datasets reveal that our suggested models

9 https://sdgs.un.org/goals.

10 https://sdgs.un.org/goals/goal1.

11 https://sdgs.un.org/goals/goal8.

12 https://sdgs.un.org/goals/goal11.

13 https://www.nytimes.com/2021/10/16/world/middleeast/iraq-sadr-election.html.

outperform other baseline techniques for the stated tasks. Using the RI metric, domain-independent TSLDA model outperformed various clustering and topic models by an average of 0.83%. Furthermore, utilizing distributed vector, we enhanced Precision to 0.89% on both R:14 and R:15 datasets. TSLDA has the advantage of eliminating the need for manually labelled lexicons or seeds by proposing Dt-WE, in which CGS was improved by using semantic similarity across the word distribution. To avoid the requirement for hand-crafted rules in implicit aspect detection, a new facet of the distributed vector is presented that is trained using TSLDA's aspect-categories. Yet, as the ABSA is a paramount field of study in the era of information explosion. Data on the internet is dramatically increasing every day, and it is raw with no class label. Manually annotating text documents is usually a tedious human task, therefore, the development of unsupervised approaches that can deal with raw data, specifically the development of nonparametric topic models like Hierarchical Dirichlet Process (HDP) could be the future path of this research.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

References

1. Najmi E, Hashmi K, Malik Z et al (2015) CAPRA: a comprehensive approach to product ranking using customer reviews. Computing 97:843–867. https://doi.org/10.1007/s00607-015-0439-8
2. Liu W, Jing W, Li Y (2020) Incorporating feature representation into BiLSTM for deceptive review detection. Computing 102:701–715. https://doi.org/10.1007/s00607-019-00763-y
3. Liu B (2017) Many facets of sentiment analysis. In: A practical guide to sentiment analysis. Socio-affective computing. Springer, Cham, pp 11–39. https://doi.org/10.1007/978-3-319-55394-8_2
4. Poria S, Chaturvedi I, Cambria E, Bisio F (2016b) Sentic LDA: improving on LDA with semantic similarity for aspect-based sentiment analysis. In: International joint conference on neural networks. IEEE, pp 4465–4473. https://doi.org/10.1109/IJCNN.2016.7727784
5. Xu X, Cheng X, Tan S et al (2013) Aspect-level opinion mining of online customer reviews. China Commun 10:25–41. https://doi.org/10.1109/CC.2013.6488828
6. Chen Z, Mukherjee A, Liu B, et al (2013) Exploiting domain knowledge in aspect extraction. In: Empirical methods in natural language processing, pp 1655–1667. https://www.aclweb.org/anthology/D13-1172.pdf
7. Blei DM, Ng AY, Jordan MI (2003) Latent Dirichlet allocation. J Mach Learn Res 3:993–1022
8. Bollegala D, Matsuo Y, Ishizuka M (2011) A web search engine-based approach to measure semantic similarity between words. IEEE Trans Knowl Data Eng. https://doi.org/10.1109/TKDE.2010.172
9. Jiang Z, Gao S, Chen L (2020) Study on text representation method based on deep learning and topic information. Computing 102:623–642. https://doi.org/10.1007/s00607-019-00755-y
10. Griffiths TL, Steyvers M (2004) Finding scientific topics. Proc Natl Conf Artif Intell 101:5228–5235. https://doi.org/10.1073/pnas.0307752101
11. Mukherjee A, Liu B (2012) Aspect extraction through semi-supervised modeling. In: Proceedings of 50th annual meeting of the association for computational linguistics, pp 339–348. https://www.aclweb.org/anthology/P12-1036
12. Andrzejewski D, Zhu X, Craven M (2009) Incorporating domain knowledge into topic modeling via Dirichlet Forest priors. In: Proceedings of 26th Annual International Conference on Machine Learning, pp 1–8. https://doi.org/10.1145/1553374.1553378
13. Hoffman MD, Blei DM, Bach F (2010) Online learning for latent Dirichlet allocation. Adv Neural Inf Process Syst 23:856–864
14. Das R, Zaheer M, Dyer C (2015) Gaussian LDA for topic models with word embeddings. In: ACL-IJCNLP 2015—53rd annual meeting association computing linguistics. 7th International joint conference on natural language processing. Asian federation on natural language processing proceedings conference. https://doi.org/10.3115/v1/p15-1077

15. Hai Z, Cong G, Chang K et al (2017) Analyzing sentiments in one go: a supervised joint topic modeling approach. IEEE Trans Knowl Data Eng 29:1172–1185. https://doi.org/10.1109/TKDE.2017.2669027
16. Lin C, He Y (2009) Joint sentiment/topic model for sentiment analysis. In: Proceeding 18th ACM conference on information and knowledge management—CIKM '09, ACM, p 375. https://doi.org/10.1145/1645953.1646003
17. 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
18. Santosh DT, Vardhan BV, Ramesh D (2016) Extracting product features from reviews using feature ontology tree applied on LDA topic clusters. In: Proceedings of 6th international conference on advanced computing conference, pp 163–168. https://doi.org/10.1109/IACC.2016.39
19. Ali F, Kwak D, Khan P et al (2019) Transportation sentiment analysis using word embedding and ontology-based topic modeling. Knowl Based Syst 174:27–42. https://doi.org/10.1016/j.knosys.2019.02.033
20. Park S-M, Lee SJ, On B-W (2020) Topic word embedding-based methods for automatically extracting main aspects from product reviews. Appl Sci 10:3831. https://doi.org/10.3390/app10113831
21. García-Pablos A, Cuadros M, Rigau G (2018) W2VLDA: almost unsupervised system for aspect based sentiment analysis. Expert Syst Appl 91:127–137. https://doi.org/10.1016/j.eswa.2017.08.049
22. Nimala K, Magesh S, Thamizh Arasan R (2018) Hash tag based topic modelling techniques for twitter by tweet aggregation strategy. J Adv Res Dyn Control Syst 10:571–578
23. Xiong S, Wang K, Ji D, Wang B (2018) A short text sentiment-topic model for product reviews. Neurocomputing 297:94–102. https://doi.org/10.1016/j.neucom.2018.02.034
24. Tang F, Fu L, Yao B, Xu W (2019) Aspect based fine-grained sentiment analysis for online reviews. Inf Sci 488:190–204. https://doi.org/10.1016/J.INS.2019.02.064
25. Tubishat M, Idris N, Abushariah MAM (2018) Implicit aspect extraction in sentiment analysis: review, taxonomy, opportunities, and open challenges. Inf Process Manag. https://doi.org/10.1016/j.ipm.2018.03.008
26. Gobi N, Rathinavelu A (2019) Analyzing cloud based reviews for product ranking using feature based clustering algorithm. Cluster Comput. https://doi.org/10.1007/s10586-018-1996-3
27. Miranda C, Buelvas E (2019) AspectSA: unsupervised system for aspect based sentiment analysis in Spanish. In: Prospectiva. https://doi.org/10.15665/rp.v17i1.1961
28. Rana TA, Cheah YN (2020) Multi-level knowledge-based approach for implicit aspect identification. Appl Intell 50:4616–4630. https://doi.org/10.1007/s10489-020-01817-x
29. Feng J, Cai S, Ma X (2019) Enhanced sentiment labeling and implicit aspect identification by integration of deep convolution neural network and sequential algorithm. Clust Comput 22:5839–5857. https://doi.org/10.1007/s10586-017-1626-5
30. Afzaal M, Usman M, Fong ACM, Fong S (2019) Multiaspect-based opinion classification model for tourist reviews. Expert Syst 36:e12371. https://doi.org/10.1111/exsy.12371
31. Mowlaei ME, Saniee Abadeh M, Keshavarz H (2020) Aspect-based sentiment analysis using adaptive aspect-based lexicons. Expert Syst Appl 148:113234. https://doi.org/10.1016/j.eswa.2020.113234
32. Xu Q, Zhu L, Dai T et al (2020) Non-negative matrix factorization for implicit aspect identification. J Ambient Intell Humaniz Comput 11:2683–2699. https://doi.org/10.1007/s12652-019-01328-9
33. Wang S, Zhou W, Jiang C (2020) A survey of word embeddings based on deep learning. Computing 102:717–740. https://doi.org/10.1007/s00607-019-00768-7
34. Almansor EH, Hussain FK, Hussain OK (2021) Supervised ensemble sentiment-based framework to measure chatbot quality of services. Computing 103:491–507. https://doi.org/10.1007/s00607-020-00863-0
35. Demeester T, Rocktäschel T, Riedel S (2016) Lifted rule injection for relation embeddings. In: Proceedings of 2016 conference on empirical methods natural language processing, pp 1389–1399
36. Rand WM (1971) Objective criteria for the evaluation of clustering methods. J Am Stat Assoc 66:846–850. https://doi.org/10.1080/01621459.1971.10482356
37. Shannon C (1948) A mathematical theory of communication | Nokia Bell Labs Journals & Magazine | IEEE Xplore. Bell Syst Tech J. https://doi.org/10.1002/j.1538-7305.1948.tb01338.x
38. Strehl A, Ghosh J (2003) Cluster ensembles—a knowledge reuse framework for combining multiple partitions. J Mach Learn Res 3:583–617
39. Poria S, Cambria E, Gelbukh A (2016) Aspect extraction for opinion mining with a deep convolutional neural network. Knowl Based Syst 108:42–49. https://doi.org/10.1016/j.knosys.2016.06.009
40. Arthur D, Vassilvitskii S (2007) K-means++: the advantages of careful seeding. In: Proceedings of 18th annual ACM-SIAM symposium discrete algorithms
41. Hoyer PO (2004) Non-negative matrix factorization with sparseness constraints. J Ambient Intell Humaniz Comput. https://doi.org/10.1007/s12652-019-01328-9

42. Dumais ST (2004) Latent semantic analysis. Annu Rev Inf Sci Technol. https://doi.org/10.1002/aris.1440380105
43. Su Q, Xu X, Guo H, et al (2008) Hidden sentiment association in Chinese web opinion mining. In: Proceedings of 17th international conference on world wide web, pp 959–968. https://doi.org/10.1145/1367497.1367627
44. Chen L, Martineau J, Cheng D, Sheth A (2016) Clustering for simultaneous extraction of aspects and features from reviews. In: 2016 Conference of the North American chapter association computing linguistics human language technologies. NAACL HLT 2016—Proceedings conference, pp 789–799. https://doi.org/10.18653/v1/N16-1093
45. Hai Z, Chang K, Kim JJ (2011) Implicit feature identification via co-occurrence association rule mining. In: Computational linguistics and intelligent text processing. CICLing Lecture notes in computer science, pp 393–404. https://doi.org/10.1007/978-3-642-19400-9_31
46. Xiong S, Ji D (2016) Exploiting flexible-constrained K-means clustering with word embedding for aspect-phrase grouping. Inf Sci (NY) 367–368:689–699. https://doi.org/10.1016/j.ins.2016.07.002
47. Xiong S, Cheng M, Batra V et al (2020) Aspect terms grouping via fusing concepts and context information. Inf Fusion 64:12–19. https://doi.org/10.1016/j.inffus.2020.06.007
48. Zhai Z, Liu B, Xu H, Jia P (2011) Clustering product features for opinion mining. In: Proceedings of 4th ACM international conference on web search data mining WSDM, vol 2011, pp 347–354. https://doi.org/10.1145/1935826.1935884
49. McAuliffe JD, Blei DM (2005) Supervised topic models. In: Advanced neural information processing systems, pp 121–128. arXiv:1003.0783
50. Jo Y, Oh A (2011) Aspect and sentiment unification model for online review analysis. In: Proceedings of 4th ACM international conference on web search data mining. ACM Press, pp 815–824. https://doi.org/10.1145/1935826.1935932
51. Wang H, Lu Y, Zhai C (2010) Latent aspect rating analysis on review text data: A rating regression approach. In: Proceedings of 16th ACM SIGKDD international conference on knowledge discovery and data mining—KDD '09. https://doi.org/10.1145/1835804.1835903
52. Chen Z, Mukherjee A, Liu B, et al (2013) Exploiting domain knowledge in aspect extraction. In: Empirical methods on natural language processing, pp 1655–1667. https://www.aclweb.org/anthology/D13-1172
53. Poria S, Cambria E, Ku L-W, et al (2015) A rule-based approach to aspect extraction from product reviews. In: Second work on natural language processing for social media, pp 28–37. https://www.aclweb.org/anthology/W14-5905.pdf
54. Schouten K, Frasincar F (2014) Implicit feature detection for sentiment analysis. In: Proceedings of 23rd international conference on world wide web, pp 367–368. https://doi.org/10.1145/2567948.2577378
55. Zeng L, Li F (2013) A classification-based approach for implicit feature identification. In: Chinese computational linguistics natural language processing based natural annotations of big data, pp 190–202. https://doi.org/10.1007/978-3-642-41491-6_18
56. Sun L, Li S, Li J, Lv J (2014) A novel context-based implicit feature extracting method. In: DSAA 2014—proceedings of 2014 IEEE international conference on data science and advanced analytics, pp 420–424. https://doi.org/10.1109/DSAA.2014.7058106
57. Yan Z, Xing M, Zhang D, Ma B (2015) EXPRS: an extended pagerank method for product feature extraction from online consumer reviews. Inf Manag 52:850–858. https://doi.org/10.1016/j.im.2015.02.002
58. Toh Z, Wang W (2014) DLIREC: aspect term extraction and term polarity classification system. In: Proceedings of the 11th international workshop on semantic evaluation (SemEval 2014), pp 235–240
59. Del-pino MA, Watine P (2020) The importance of poverty in sustainability policies: an approach to understanding online opinion. In: Actas del III Congreso Internacional de Ingeniería de Sistemas La, pp 183–194

Omar Mustafa AL-Janabi is a PhD candidate at university of Science Malaysia (USM). Received his B.S. and M.S degree in Computer Science from University of Al-Ma'mun, and Acharya Nagarjuna University (ANU), respectively. His research focuses on data science, Bayesian probabilistic models, sentiment analysis, natural language processing and deep Bayesian models.

Nurul Hashimah Ahamed Hassain Malim received her B.Sc (Hons) & M.Sc Computer Science from Universiti Sains Malaysia, Malaysia. She obtained her PhD from The University of Sheffield, United Kingdom. Her current research interests include High Performance Computing, Chemoinformatics, Bioinformatics, Data Analytics and Sentiment Analysis. She is currently a Senior Lecturer in the School of Computer Sciences, Universiti Sains Malaysia.

Yu-N Cheah received his B.Sc (Hons) Computer Science and completed his PhD in Computer Science from Universiti Sains Malaysia, Malaysia. His current research interests include Knowledge Management and health informatics which specialized in knowledge management, knowledge engineering, intelligent systems, semantic technologies & health informatics. He is currently a Senior Lecturer in the School of Computer Sciences, Universiti Sains Malaysia.