ISBN: 978-81-958975-8-2
https://doi.org/10.5281/zenodo.20770812

Chapter 1

SAR-HDP: Non-parametric Topic Model for Aspect categorisation based on online reviews

Omar Mustafa AL-Janabi[1*], Nurul Hashimah Ahamed Hassain Malim[2], Cheah Yu-N[3], Osamah Mohammed Alyasiri[4], Aseel Musa Jasim[5]

[1]IT Division, College of Medicine, University of Baghdad, Baghdad, Iraq.
[2]School of Computer Sciences, Universiti Sains Malaysia, Penang, Malaysia.
[3]School of Computer Sciences, Universiti Sains Malaysia, Penang, Malaysia.
[4]Karbala Technical Institute, Al-Furat Al-Awsat Technical University, Karbala, Iraq.
[5]Department of Public Administration, College of Administration and Economics, University of Baghdad, Baghdad, Iraq.
*Corresponding Author.

Abstract:

Aspect categorisation and its utmost importance in the field of Aspect- based Sentiment Analysis (ABSA) has encouraged researchers to improve topic model performance for modelling the aspects into categories. In general, a major- ity of its current methods implement parametric models requiring a pre-determined number of topics beforehand. However, this is not efficiently undertaken with unan- notated text data as they lack any class label. Therefore, the current work pre- sented a novel non-parametric model drawing a number of topics based on the semantic association present between opinion-targets (i.e., aspects) and their re- spective expressed sentiments. The model incorporated the Semantic Association Rules (SAR) into the Hierarchical Dirichlet Process (HDP), named ('SAR-HDP'). The phrase-based (or aspect-based) Bayesian model (SAR-HDP) did not consider the word's sentence being drawn from a single topic due to the presence of mul- tiple aspects in a single review, which belonged to a multiple-aspect topic (i.e., category). Beyond its consideration of the semantic information for aspect iden- tification, the proposed model further upheld the semantic information discerned between the drawn topics and aspects identified to maintain topic consistency. Empirical investigation showed that the approach positioned successfully outper- formed standard parametric models and nonparametric models in terms of aspect categorisation when subjected to restaurant and hotel reviews sourced from Ama- zon and TripAdvisor.

Keywords : Non-parametric models. Hierarchical Dirichlet process. Collapsed Gibbssampling Aspect extraction. Aspect categorisation. Online reviews

ISBN: 978-81-958975-8-2

1. Introduction

The exponential growth of Social Web services has expanded the amount and breadth of user-generated content generally found on the World Wide Web (WWW).Such content came from people's ideas and opinions, yielding an excellent oppor- tunity for the respective industries to capture customer satisfaction regarding their products offered, as well as helping the public in deciding on their products by implementing feature ranking applications [1]. To ensure the practicality of these contents, ABSA models have emerged; they are tasked with extracting essential figures of speech in opinionated customer reviews, which is called an aspect (or aspect-term) to determine its expressed sentiment [2].

Over the last decade, a notable number of researchers have presented different methods for extracting aspects across different domains and languages [3]. They are thus classifiable as either unsupervised [4–10], semi-supervised [11–15], or su- pervised methods [16–19]. In terms of the ABSA subtask, its methods are primarily concerned with fusing the syntactical relation for aspect extraction. Furthermore, three different approaches have been previously implemented to extract aspects; the first approach is an augmented frequency-based method proposed to utilise semantic relatedness [4, 5, 20–24]. The second approach is the dependency-based [9, 25–29] and third one denotes using pattern-based methods [7, 30–33] built based on the syntactic relation between aspect-terms and its related opinion-words. Typically, extraction types based on the augmented frequency-based, dependency- based, and pattern-based approaches generate promising results in certain scenar- ios. However, augmented frequency-based approaches have been known to ignore the infrequent aspect terms as they mostly retrieve the high-frequency nouns and noun-phrases as the aspects. Meanwhile, the dependency-based or pattern-based approaches are restricted by the domain dependency, wherein the rules are thus ar- ticulated according to a particular domain or language [21-24]. Moreover, neither of any aspect extraction approaches mentioned above can be employed in fur- ther identifying the aspect category for aspect-terms perceived, nor the associated opinion-word/sentiment.

Alternatively, probabilistic topic models are utilised in data mining and la- tent data discovery, as well as its most important implementation for discerning the relationship between data and text documents. Therefore, researchers have implemented the method to conduct aspect analyses [34, 35]. For example, La- tent Dirichlet Allocation (LDA) [36] is a popular unsupervised admixture or topic model, which specifically builds a set of topics based on an input collection of documents. As a parametric model, it incorporates the concept of Bag of Words (BoWs) for the articulation of topics; the standard LDA has been adopted by many researchers for aspect extraction [34, 35, 37, 38]. Meanwhile, others have considered lexicon-based topic models [39–41], whereas works such as [42–45] have opted for distributed vector method and knowledge-based information as a supplementary approach

for aspect analysis. Regardless, a gap in the current studies has been underlined in which the existing probabilistic methods are not only paramet- ric models requiring a class label; they are also seemingly dismissive about finding the semantic association between the aspect-terms and the expressed sentiments for each aspect topic (category).

As a solution for the lack of semantic regularities in identifying aspect-terms, a SAR algorithm is developed. It encompasses the semantic regularities towards identifying the aspect terms and their expressed sentiment while working in tan- dem with a non-parametric model, that is HDP. Here, HDP builds on the premise of the Multivariate Gaussian Distribution (MGD) in modelling the identified aspects into their respective categories. Then, the hyperparameter values are up- dated by using a sampling algorithm (i.e. Gamma Distribution).

2.Problem Definition of Topic Models on Online Reviews

To date, traditional topic models such as LDA are widely implemented in ABSA subtasks, such as aspect-categorisation. However, they also come with flaws:

- Ignore the word order.
- Ignore semantic regularities.

Even though these two limitations are being tackled by utilising lexical supple- mentary approaches like lexicon-based methods and distributed vectors [39–42], they remain limited. This is attributable to the reliance of presented ABSA meth- ods on an annotated number of topics, which is not always the case in the real world. Besides, a class label is absent in online reviews (“text data”) found on pub- lic platforms like Amazon or TripAdvisor, which leaves the current methods lack- ing. The best example for this is the online reviews in SemEval-2014 (Restaurant and Laptop) domain data. Here, the restaurant corpus (Fig. 1) is annotated into five aspect categories, including “Price” and “Food”, following which the anno- tated aspect term in the opinionated sentence is further assigned to an aspect topic (category). In contrast, Fig. 2 reveals the lack of aspect topic (“aspect-category”) for the annotated aspect-terms in the Laptop corpus (“SemEval-2014”). There- fore, this work introduces a novel non-parametric Bayesian model that addresses three subtasks simultaneously:

- Detects the aspect and its expressed sentiment.
- Detects aspect categories automatically.
- Allocate the aspect terms into categories.

```
<sentence id="813">
    <text>All the appetizers and salads were fabulous, the steak was mouth
watering and the pasta was delicious!!!</text>
    <aspectTerms>
      <aspectTerm term="appetizers" polarity="positive" from="8" to="18"/>
      <aspectTerm term="salads" polarity="positive" from="23" to="29"/>
      <aspectTerm term="steak" polarity="positive" from="49" to="54"/>
      <aspectTerm term="pasta" polarity="positive" from="82" to="87"/>
    </aspectTerms>
    <aspectCategories>
      <aspectCategory category="food" polarity="positive"/>
    </aspectCategories>
  </sentence>
```

Figure 1: Annotated aspect category in the SemEval-2014 restaurant domain data

```
<sentence id="2339">
    <text>I charge it at night and skip taking the cord with me because of the
good battery life.</text>
    <aspectTerms>
      <aspectTerm term="cord" polarity="neutral" from="41" to="45"/>
      <aspectTerm term="battery life" polarity="positive" from="74"
to="86"/>
    </aspectTerms>
  </sentence>
```

Figure 2 : Unannotated aspect-categories in the SemEval-2014 laptop domain data

3. Research Objectives

The objectives of the proposed model differ from currently available methods, wherein it does the following accordingly:

1. Develops semantic association rules (SAR) for the detection of aspect-terms and their expressed sentiments.
2. Employs word embeddings (WE) to maintain semantic regularities on topics.
3. Automatically determine the number of aspects and their topics based on the developed SAR, which works in tandem with the HDP model.
4. Updates the hyperparameter values by using a sampling algorithm.

4. Recent Work

Topic models are generally differentiated into the unsupervised machine learning methods. An advantage of employing topic models as opposed to other methods is its implementation in handling multiple subtasks simultaneously in ABSA, in- cluding aspect extraction, and categorisation [35]. Currently, the proposed topic models for aspect-analysis are mainly built on the premises of parametric model (i.e. LDA) as seen in different surveys such as [34, 35, 46], whereas a few others rely on its non-parametric counterpart (i.e. HDP) [47, 48]. Along with the topic models that are presented for aspect categorisation, other methods (e.g., clustering methods) are briefly stated in the following subsections.

4.1 Parametric model

This section includes some of the recent topic models that have performed aspect modelling by using labelled datasets ("that have a specific number of topics"). For example, the enhanced topic model "Sentic LDA" comprises a lexicon-based method (i.e., SenticNet) for semantic consideration, whereby its evaluation is car- ried out using a specified number of topics sourced from the hotel reviews dataset.The dataset is annotated into seven aspect categories (or topics) prior to the as- sessment process [39]. Meanwhile, the AEP-LDA [49] model operates under the assumption that all words found in a sentence are drawn from one topic. This is not a true assumption, however, due to the presence of multiple aspects within a single sentence belonging to a different topic(s).

In [50], semantic regularities have been presented by using word embeddings. However, the model was not tasked with modelling the aspect terms into categories based on the identified aspect; instead, a clustering algorithm was utilised for this purpose. Similarly, W2VLDA [42] has been highlighted in articulating aspect terms into their respective categories in consideration of semantic regularities. Regardless, it is a parametric model and thus requires a class label. The main problem with the existing topic models (e.g., LDA) lies in the manner in which the proposed models presume the document is assigned to a single topic. This is not always true as it may be equipped with multiple aspect topics.

4.2 Non-parametric models

Non-parametric models are introduced to model unannotated aspect-categories in which their evaluation is carried out using an annotated counterpart. In line with this, scholars such as Ding et al. [47] and Yang et al. [51] have performed aspect- sentiment identification by utilising a hybridised topic model, which relies on LDA and HDP models.

4.3 Other Methods for Aspect Categorisation

Dictionary-based methods being used to address the aspect categorization [52, 53]. It is seemingly limited due to the development of new vocabulary. Others used clustering algorithms [54–57] for aspect categorization. Using the clustering algorithms for aspect categorization is deficient for unlabelled data and it does not consider semantic regularities between the clusters.

5. Proposed Framework

This work presented a novel framework tasked with addressing the major chal- lenges in determining the association between aspect and sentiment words seman- tically. Then, it concurrently allocates the identified aspects into their respective aspect topic (category) accordingly as shown in Fig. 4. Therefore, Subsection (5.1) depicts the SAR for aspect and sentiment determination in the reviews, while Subsection (5.2) displays the manner in which it works in tandem with the HDP model.

5.1 Semantic Association Rules (SAR)

In ABSA, a syntactical relation can be perceived between an aspect and its senti- ment words [37, 42, 49]. Recent topic model methods (e.g., W2VLDA, AEP-LDA,HDP-LDA, and SenticLDA), however, seemingly perceive the sentence-level (sen- tence review) to hold a single aspect, which is not always true. This is reflected in the aforementioned example (Fig. 1) in which ‘appetisers’ and ‘salads’ belong to the aspect topic ‘food’ despite them being multiple aspect-terms from a sin- gle sentence. Therefore, this work proposed the SAR algorithm to determine the aspect and its sentiments at the phrase level.

For each phrase p in an opinionated sentence s of a set of sentences Sn = {s1, s2, s3, si}, Sn is a set of all opinionated sentence reviews (e.g., s1) in the cor- pus. Here, the SAR would determine whether the drawn phrase p is an aspect, sentiment, or objective word. One may note that the aspect words are usually nouns or noun phrases, while they may be adjectives if referring to the sentiment words. Alternatively, objective words are unrelated to the aspect-based analysis altogether. As an algorithm, SAR generally leverages the semantic association to be combined with the syntactic rules (Table 1), wherein the associations between the aspects and their expressed sentiments are attained by using the word embed- dings [58].

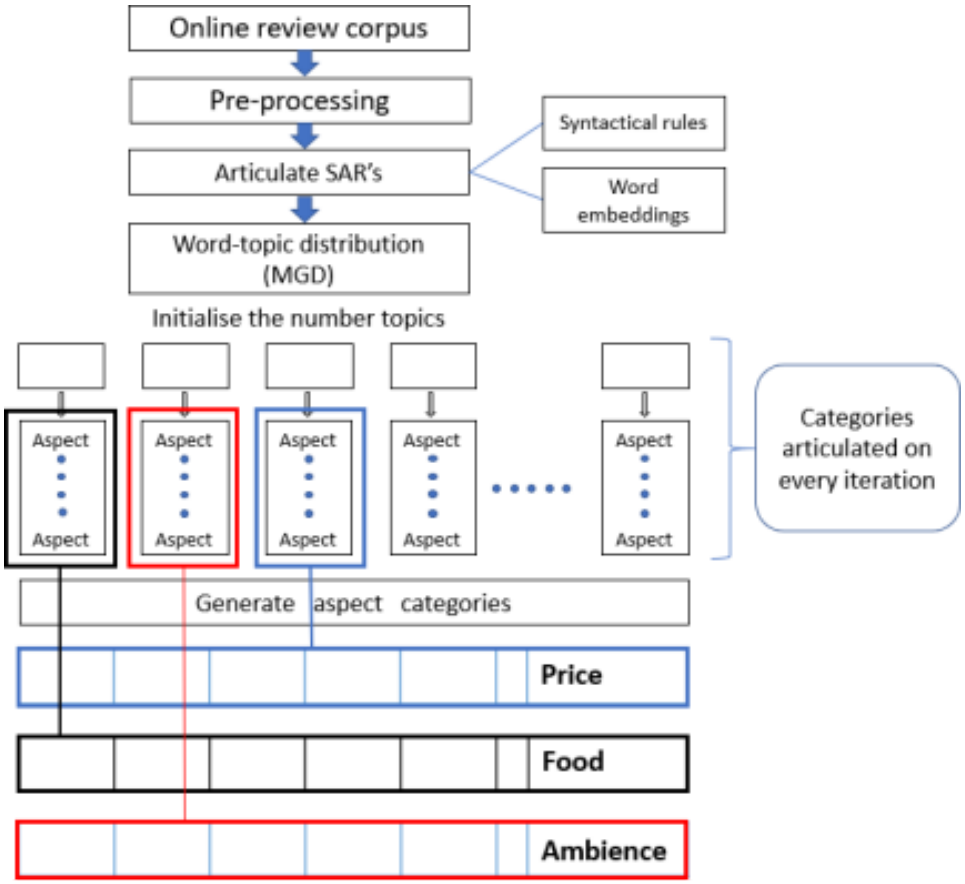

Figure 3 : Semantic Association Rules-Hierarchical Dirichlet Process (SAR-HDP) Framework

First, the common syntactical relation between the aspects and sentiments were identified by employing the formulated rules. For instance, 'NN' and 'JJ' are used to extract the aspect of 'picture quality' (I) (Table 2).

However, extraction by using the syntactic rules detailed in Table 1 is not always correct as the 'free speakerphone' is a meaningless aspect term extracted by using the same syntactic rule. Therefore, the novelty of this subtask is rooted in the introduction of semantic association to eliminate any irrelevant and improperly extracted aspects. Here, soft-cosine [59] is depicted accordingly in ascertaining the correlation between the words as per the Vector Space Model (VSM). Unlike other similarity measures, the soft-cosine measure can be trained by using prior knowledge sourced from the word embeddings [46]. As a result, measurement of the semantic association scale is undertaken based on the prior knowledge fed into the similarity measure.

Table 1 Articulated Syntactic Rules

Patterns	Description		
NN → JJ	Noun, common, singular, mass	→	Adjective or numeral, ordinal
NNS → JJ	Noun, common, plural	→	Adjective or numeral, ordinal
NNP → JJ	Noun, proper, singular	→	Adjective or numeral, ordinal
NNPS → JJ	Noun, proper, plural	→	Adjective or numeral, ordinal
NN → JJR	Noun, common, singular, mass	→	Adjective, comparative
NNS → JJR	Noun, common, plural	→	Adjective, comparative
NNP → JJR	Noun, proper, singular	→	Adjective, comparative
NNPS → JJR	Noun, proper, plural	→	Adjective, comparative
NN → JJS	Noun, common, singular, mass	→	Adjective, superlative
NNS → JJS	Noun, common, plural	→	Adjective, superlative
NNP → JJS	Noun, proper, singular	→	Adjective, superlative
NNPS → JJS	Noun, proper, plural	→	Adjective, superlative
JJ ← NN	Adjective or numeral, ordinal	←	Noun, common, singular, mass
JJ ← NNS	Adjective or numeral, ordinal	←	Noun, common, plural
JJ ← NNP	Adjective or numeral, ordinal	←	Noun, proper, singular
JJ ← NNPS	Adjective or numeral, ordinal	←	Noun, proper, plural
JJR ← NN	Adjective, comparative	←	Noun, common, singular, mass
JJR ← NNS	Adjective, comparative	←	Noun, common, plural
JJR ← NNP	Adjective, comparative	←	Noun, proper, singular
JJR ← NNPS	Adjective, comparative	←	Noun, proper, plural
JJS ← NN	Adjective, superlative	←	Noun, common, singular, mass
JJS ← NNS	Adjective, superlative	←	Noun, common, plural
JJS ← NNP	Adjective, superlative	←	Noun, proper, singular
JJS ← NNPS	Adjective, superlative	←	Noun, proper, plural

<sentence id=”1480”> <text>My friend reports the notebook is astonishing in performance, picture quality, and ease of use.</text> (I)

Note that the opinion target terms ‘picture’ and ‘quality’ from sentence (I) are different terminologies; hypothetically, they should be mapped to different dimen- sions in the VSM despite being semantically related. To this end, word embeddings denote the notion that the relatedness of two words relies upon their inherent context in the reviews. If they are frequently occurring in the same context, they are henceforth semantically related.

In practice, implementing syntactic rules from Table 1 on the review sentence
(I) results in each noun ‘NN’ being associated with the adjective ‘JJ’ to which it is nearest in which two texts are identified, namely: (1) picture, and (2) quality. Similarity measure defines two dimensions in the VSM by using the following feature terms: picture and quality, as represented by two vectors, α and β:

$$\alpha = [1, 0]$$

$$\beta = [0, 1]$$

The traditional cosine similarity yields the '0' value as a similarity score between the two vectors. However, upon consideration of the semantic similarity for the words based on the prior knowledge of domain-trained word embedding, a com- parably high similarity is generated in the VSM for such vectors.

To transfer the data vectors into a space of dimension N^2, the data vectors $a = (a_i)$, $b = (b_i)$ to a new N^2 *dimensional* vectors (a_{ij}), (b_{ij}) by averaging different coordinates:

$$a_{ij} = \sqrt{s_{ij}}\frac{a_i + a_j}{2}, \qquad b_{ij} = \sqrt{s_{ij}}\frac{b_i + b_j}{2}, \tag{1}$$

Where $s_{ij} = sim(f_i, f_j)$; thus, it is a natural assumption that the similarity is modelled as cosine between these two objects:

$$cosine(e^i, e^j) = s_{ij} = sim(f_i, f_j) \tag{2}$$

Given two vectors (*i.e.* e^i, e^j), cosine(e^i, e^j) calculates the cosine similarity measure between them, which is the normalised dot-product of the two vectors. Besides, f_i and f_j are the vectors corresponding to (a_{ij}) *and* (b_{ij}), respectively, while $sim(.)$ is the semantic similarity. To this end, the data points or words compared are calculated by using soft_cosine (s_c):

$$s\_c(a,b) = \frac{\sum\sum_{i,\,j=1}^{N} a_{ij}b_{ij}}{\sqrt{\sum\sum_{i\,,\;j=1}^{N} a_{i,\;j}^2}\;\sqrt{\sum\sum_{i,\;j=1}^{N} b_{ij}^2}} \tag{3}$$

Here, the straightforward interpretation of Formula (3) considers the similar- ity held by a pair of features or opinion target words concerning their semantic contexts.

Besides, a threshold value is set to determine the semantic similarity between the opinion target and opinion words. If the similarity measure between 'NN' and 'JJ' for picture quality in the sentence (I) is less than the predetermined threshold, the opinion target will be ignored. For example, if the calculated similarity of the soft-cosine measure is less than 0.2 when the determined threshold is between 0 to 1, the opinion target candidates are thus most likely to be non-existent opinion targets.

Alternatively, algorithm (1) demonstrates the proposed SAR algorithm in iden- tifying the opinion targets by utilising the modified syntax-based method and semantic association. The method implementation is possible due to the two de- fined vectors, namely a_{ij}, and b_{ij}. Accordingly, the pair of opinion target words

is calculated in terms of their similarity, whereby the pair is taken from the prela- belled syntactic rules. For example, the aspect term is represented as “picture” in comparison with the following opinion word (i.e., “quality”) according to the syn- tactic rule (1). Thus, the semantic similarity threshold will further determine the similarity relatedness of the noun “picture” and adjective “quality” accordingly.

Besides, Table 2 depicts the extracted opinion target based on the stated syn- tactic rules. The patterns that have “NN” and are followed by “JJ” are obtained as follows: if the noun is semantically associated with the subsequent adjective and the noun extracted as an aspect term along with it is a sentiment word. SAR is fused into the drawing process of word-topic distribution (i.e., MGD) in the proposed non-parametric model (i.e., HDP) to model the identified aspect into category(s) in tandem as shown in the following section.

Table 2 The Identified Aspect

Opinion targets	Syntactic rule	Sentence
Picture	(1)	My friend reports the notebook is astonishing in performance, (JJ) picture (NN) quality, and ease of use.

Algorithm 1 : Semantic Association Rules (SAR)

```
Opinion-word ← [ ]
    Opinion-word can be {'JJ','JJR','JJS'}
Opinion-target ← [ ]
    Opinion-target can be {'NN','NNS','NNP','NNPS'}
Semantic-similarity ← [ ]
For each sentence s in sentences Sn:
    s ← part of speech tags
For each tagged sentence s:
    Apply syntactic rules (Table (1))
For each Opinion-word in the tagged reviews, find the nearest Opinion-target:
    For a in Opinion-word:
        For b in Opinion-target:
            Semantic-similarity = s_c(a, b)
            If float(Semantic-similarity) < threshold:
                Pass
            Else:
                If float(Semantic-similarity) ≥ threshold:
                    Opinion-target = b
```

5.2 SAR-HDP MODEL

The central ideation of the proposed non-parametric model (i.e., SAR-HDP) is towards accommodating the level of complexity embedded in the data in line with the graphical model shown in Fig. 5. Nevertheless, the challenge perceived is in carrying out joint mining for the aspect topic (i.e., category) and aspect terms within a single sentence review. This is attributable to the possibility of multiple aspects belonging to either the same or different aspect topic (category). As these reviews are unannotated, the supervised methods are rendered lacking due to their requirement for large-scale annotated data. Therefore, unlike the fundamental as- sumption underlying the HDP, which denotes that the words in one document are exchangeable [60, 61], the SAR-HDP model (algorithm 2) indicates that the topic assignments of words are conditionally dependent. The process is thus underpinned by the semantic association between the identified 'aspect' by using ASR and the aspect topic of 'category'.

To this end, the grouping of aspect terms p identified into aspect topic K could be carried out accordingly (e.g., aspect terms 'appetisers', 'salads', and 'steak' being grouped into aspect topic "food" as in Fig. 1). Here, the MGD proposes to draw aspect topic K in the ASR-HDP model in consideration of the embedding space R^M with expectation μ_k and covariance matrix Σ_k. In practice, the sentence review S comprises phrases P and the phrase p is identified as aspects by using ASR, which are then represented using word embeddings symbolised by $v(p) \in R^M$. The embedding of phrase p (or $v_{d,i}$) indexes a vector in the document d at position i, with M dimension (aka window size) in the embedding space R^M. This signifies the selection of a phrase vector from topic k drawn from $N(\mu_k, \Sigma_k)$. Therefore, the conjugate priors of μ_k *and* Σ_k are:

First: covariance value sampled using the Inverse Wishart Distribution (IWD) $\sum_k \sim W^{-1}(\psi, v)$. Σ_k is presented to follow the IWD sampled as $W^{-1}(\psi, v)$. IWD is a probability distribution defined for $\psi \geq 0$ that is a real-valued positive-definite matrix of size M * M, whereby $\psi = 3.I_{M*M}$.

Where v is the degree of freedom that is greater than M (i.e., $v = M + 1$).

Where the IWD equals the length of M (i.e., $W^{-1} = len(M)$).

Second: covariance Σ_k sampled using a zero-centred Gaussian distribution for the mean $\mu_k \sim (0, \frac{\Sigma_k}{k})$. Consequently, the similarity distance can be found between the two vectors of μ_k and Σ_k by using the cosine measure.

Table 3 Notations

Key	Description
k	The initial number of topics
α	Parameter of topics prior
γ	Parameter of tables prior
V	Number of unique phrases in the vocabulary
μ_k	Mean
Σ_k	Covariance
Doc	List of documents
$\bar{\tau}$	Prior proportion
$\bar{\vartheta}_k$	Multinomial parameter of a document to topic distribution
M	Number of documents

DP can be interpreted in two levels of distribution: Base level (or Base distribution) of the $\bar{\tau}$ is a **Dirichlet process** (DP) parameter, which draws the number of topics k, $\bar{\tau} \sim Dir\left(\frac{\gamma}{k}\right)$. The distribution of the topics is modelled to an infinite number of topics $K \rightarrow \infty$ and the descendant distribution is obtained given documents sampled using the multinomial distribution ϑ_m sampled as $\vartheta_m \sim Multi\left(\alpha\bar{\tau}\right)$ (see Fig. 5). $\bar{\tau}$ is the prior proportion sampled based on the new components created from the document to topic distribution, which is otherwise termed as $n_{m,k}$, and then drawn from DP. This is done using the Chinese Restaurant Process (CRP), namely a sequence of Bernoulli trials for the number of Documents and number of Topics k as shown in Equation (4).

$$p\left(m_{m,k,r}=1\right)\frac{\alpha\tau_k}{\alpha\tau_k+r-1}\forall r\in[1,n_{m,k}],m\in[1,M],k\in[1,K] \tag{4}$$

Hence, the posterior sample is obtained via:

$$\bar{\tau} \sim Dir\left(\{m_k\}k,\gamma\right) \; while \; m_k = \sum_m\sum_r m_{m,k,r} \tag{5}$$

Accordingly, the posterior sample of the base distribution is updated while a revision of the second-level HDP (i.e., descendant distribution) is generated. Meanwhile, the prior parameter τ is simulated according to the number of tables and customers as shown in Equation (6):

$$p(m_{m,k}|\alpha,\; n_{m,k}) \propto s(n_{m,k},m_{m,k})(\alpha\tau_k)^m \tag{6}$$

Where $s\left(n,m\right)$ is an unsigned Stirling number of the permutation cycles of size m in a set of n elements.

Algorithm 2 : SAR-HDP Model

1: $\alpha \sim Gam(\alpha_\alpha + T - \sum_m u_m,\ b_\alpha - \sum_m log\ v_m)$
2: $\gamma \sim Gam(\alpha_\gamma + K - 1 + u,\ \alpha_\gamma - log\ v)$
3: $\bar{\tau} \sim CRP(\gamma/k)$
4: $\vartheta_m \sim Dir(\alpha\ \bar{\tau})$
5: $Z_{m.n} \sim Cat(\vartheta_m)$
6: $P_{m,n} \sim N(m_{z,n},\ \Sigma_{z,n})$
7: $P_{m,n}$ distribution conditioned on the SAR:
$P_{m,n} \sim (N(m_{z,n},\ \Sigma_{z,n})\ if\ \ n\ \ is\ \ opion - target$

Inference - Since the word-to-topic distribution is intractable, it cannot be performed by using normal or multinomial distribution. Thus, it is either sampled via optimisation (e.g., variational inference) or sampling algorithm (e.g., Collapsed Gibbs Sampling (CGS)). CGS is duly introduced to condition the aspects to topics in the case of the current work. Regardless, the following equation represents the conditional distribution of an initial number of topics that integrate out priors' parameters. Table 3 states the used notations in the proposed non-parametric model.

$$p\left(z_i = k|.\right) \propto \left(n_{m,k}^{\neg i} + \alpha\tau_{k+\frac{1}{V}}\right) * t_{vk} - M + 1\ \left(v_{d,i} \mid \mu_k,\ \frac{k_k + 1}{k_k}\ \Sigma_k\right) \tag{7}$$

Alternatively, Equation (7) shows α as a scaler parameter, where the latent topics are $K + 1$, as it is depicted in a non-parametric model (Fig. 5). Here, the proportion is $\alpha\tau_{k+\frac{1}{V}}$, which represents all unseen components. Whenever a new word is sampled, a new topic will be created. In the second part of the equation, t_{vk} is the multivariate $t-$ distribution with v degrees of freedom and k topics.

Hyperparameter Sampling – priors are placed on two levels of concentration. Here, the sampling of u, v is pursued by using Bernoulli and Beta distributions both as shown in Equation (8), whereas γ is sampled by implementing Gamma distribution as depicted in Equation (9).

$$u \sim Bern\left(\frac{T}{T + \gamma}\right),\ v \sim Beta\left(\gamma + 1,\ T\right) \tag{8}$$

$$\gamma \sim Gam\left(\alpha_\gamma + K - 1 + u,\ b_\gamma - logv\ \right) \tag{9}$$

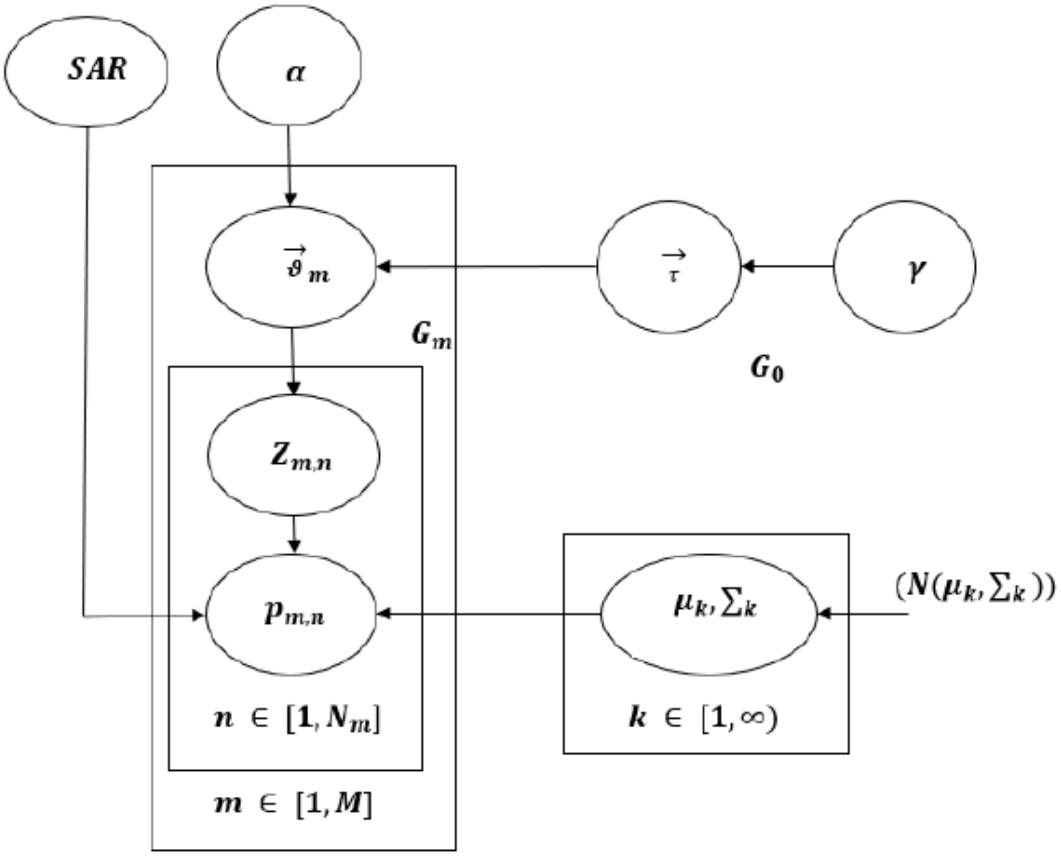

Figure : 4 Semantic Association Rules-Hierarchical Dirichlet Process (SAR-HDP) graphical model

$T = \sum_k m_k$ is the total number of tables in CRP serving as the words to be drawn from the base distribution. In contrast, the distribution of second-level CRP by using u_m , v_m is sampled as follows:

$$u_m \sim Bern\left(\frac{n_m}{n_m + \alpha}\right), v_m \sim Beta\left(\alpha + 1, n_m\right) \tag{10}$$

$$\alpha \sim Gam\left(\alpha_\alpha + T - \sum_m u_m,\ b_\alpha - \sum_m logv_m\right) \tag{11}$$

6. Experimental analysis and evaluation

This section presents a discourse of the experimental analysis and evaluation car- ried out for the proposed model, which is then compared to the results obtained by implementing related techniques.

6.1 Datasets and evaluation metrics

Datasets and evaluation metrics are vital components in assessing the building model performance. Table 4 presents the four utilised datasets to assess the SAR- HDP model's performance in context of aspect categorisation. As an assessment metric, however, the Rand Index (RI) [62], Entropy [63], Normalized Mu- tual Information (NMI) [64], and Purity [62] are utilised to measure the model efficiency upon its provision for the purpose of aspect categorisation. The index measures the similarity between the aspect-category of the proposed model and the "ground-truth" aspect-category discerned in the dataset. For example, TripAdvi- sor: Hotel dataset (TA:H) is an instance of the ground-truth data and encompasses seven categories (i.e. rooms, cleanliness, value, service, location, check-in, and business).

Table 4 Dataset Characteristics

Dataset	Short-form	#Category	#Text Granularity	#Reviews
TripAdvisor: Hotel	"TA:H"	7	Document-level	587095
SemEval-2014 Restaurant	"R:14"	5	Sentence-level	3041
SemEval-2015 Restaurant	"R:15"	5	Sentence-level	1315
SemEval-2016 Restaurant	"R:16"	6	Sentence-level	2000

6.2 Experimental setup and results

The experimental design of the proposed model articulated based on the proposed SAR-HDP model. The proposed model is a non-parametric Bayesian model built on the premises of the HDP model. The experimental setup of the SAR-HDP model dealt with two different settings:

1. SAR-HDP(CRP): the DP in this setting is interpreted using CRP as presented in section 5.2.

2. SAR-HDP(SBP): the DP in this setting is interpreted using Stick-breaking Process (SBP) as in [66].

For these settings, there are number of common and distinct hyperparameters which need to be tuned. The common hyperparameters of SAR-HDP(CRP) and SAR-HDP(SBP) included the following: α (base-level distribution), γ (descendant distribution), and the number of iterations. While the distinct hyperparameters are belongs to SBP which are indexed by k = 1, 2, 3, . . . , and β, where k is the number of weights (i.e., categories) and β is 1.5. Accordingly, Table 5 displays the values of hyperparameters for each dataset, wherein their nominated values for SAR-HDP are inspired by some of the previously proposed non-parametric Bayesian models [43, 67–70]. Since each parameter necessitated careful set-up to maintain a high level of accuracy, the optimal values were set based on the sampling algorithm.

Table 5 Parameter Values For The Proposed Non-parametric Model

	Parameters		
Datasets	α	γ	**Iterations**
R:14	0.05	1.5	2000
R:15	0.05	1.5	2000
R:15	0.05	1.5	2000
TA:H	0.05	2	2000

The central ideation of the SAR-HDP is towards accommodating the level of complexity embedded in the data. Nevertheless, the challenge perceived is in carrying out joint mining for the aspect topic (i.e., category) and aspect terms within a single sentence review. This is attributable to it possibly having multiple aspects belonging to either the same or different aspect topic (category). As these reviews are unannotated, the supervised methods are rendered lacking due to their requirement for large-scale annotated data. Therefore, unlike the fundamental assumption underlying the HDP which denotes that the words in one document are exchangeable [49,50], the SAR-HDP model indicates that the topic assignment of words are conditionally dependent. Therefore, the process is underpinned by the semantic association between the identified ‘aspect’ and the aspect topic of ‘category’. To this end, grouping the identified aspect terms into aspect topic (e.g., aspect terms “appetisers”, “salads”, and “steak” being grouped into aspect topic “food”) can be carried out accordingly.

Here, the MGD proposed drawing the aspect topic K in the SAR-HDP model in consideration of the embedding space R^M with expectation μ_k and covariance matrix Σ_k as detailed in section (5.2). To semantically guide the distribution in SAR-HDP model, the identified aspect terms in the data is represented by a vector in a multi-dimensional space, named word vector (i.e., Word2Vec [1]).

Table 6 reports the comparative analysis entitled a comparison of the model performance according to the coverage of the aspect-terms in each dataset. The proposed SAR-HDP is a generative model that being used to perform the aspect categorisation by first initialize the topics with a randomly selected aspect terms, then the topics being rectified based on the SAR and semantic regularities, so that, after a generative process of the model. It will generate number of topics that comprise semantically related aspects in each topic and ignore the non-aspects or objective words. The SAR-HDP model yielded results comparable to those obtained via the current methods (e.g., clustering and non-parametric methods) by using the RI, NMI, Entropy, and Purity evaluation measures. As shown in Figs. 5,6,7, and 8, the performance of utilising CRP as a DP distribution on the proposed non-parametric model outperformed the performance of using SBP on the investigated datasets.

Table 6 Compare the RI, NMI, Entropy, and Purity score of different configurations for SAR-HDP in terms of topic interpretation (i.e., SBP, CRP)

Configuration	R:14	R:15	R:16	TA:H
	RI			
SAR-HDP(CRP)	0.61	0.59	0.63	0.75
SAR-HDP(SBP)	0.77	0.71	0.75	0.83
	NMI			
SAR-HDP(CRP)	0.33	0.25	0.29	0.4
SAR-HDP(SBP)	0.4	0.31	0.34	0.52
	Entropy			
SAR-HDP(CRP)	1.92	1.9	1.92	1.69
SAR-HDP(SBP)	1.69	1.69	1.65	1.64
	Purity			
SAR-HDP(CRP)	0.4	0.41	0.41	0.61
SAR-HDP(SBP)	0.62	0.65	0.63	0.67

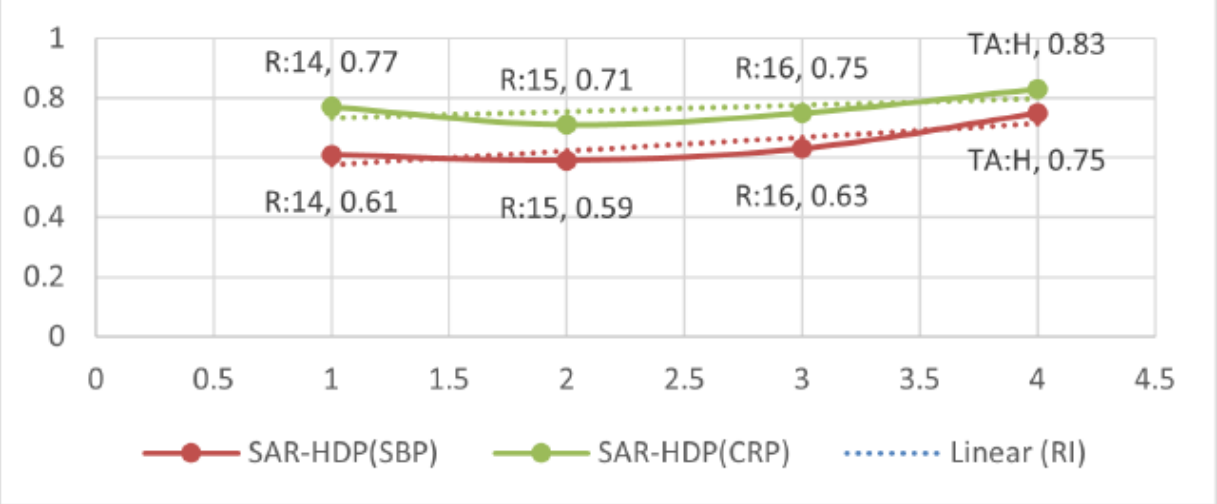

Figure : 5 Experimental results with the variation SAR-HDP settings (i.e., SBP, CRP) in terms of RI score

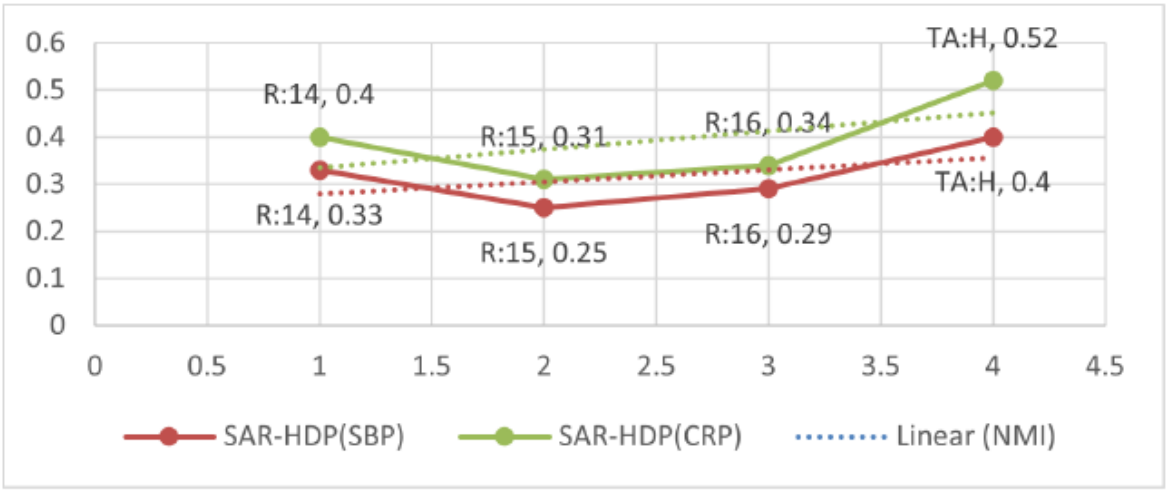

Figure : 6 Experimental results with the variation SAR-HDP settings (i.e., SBP, CRP) in terms of NMI score

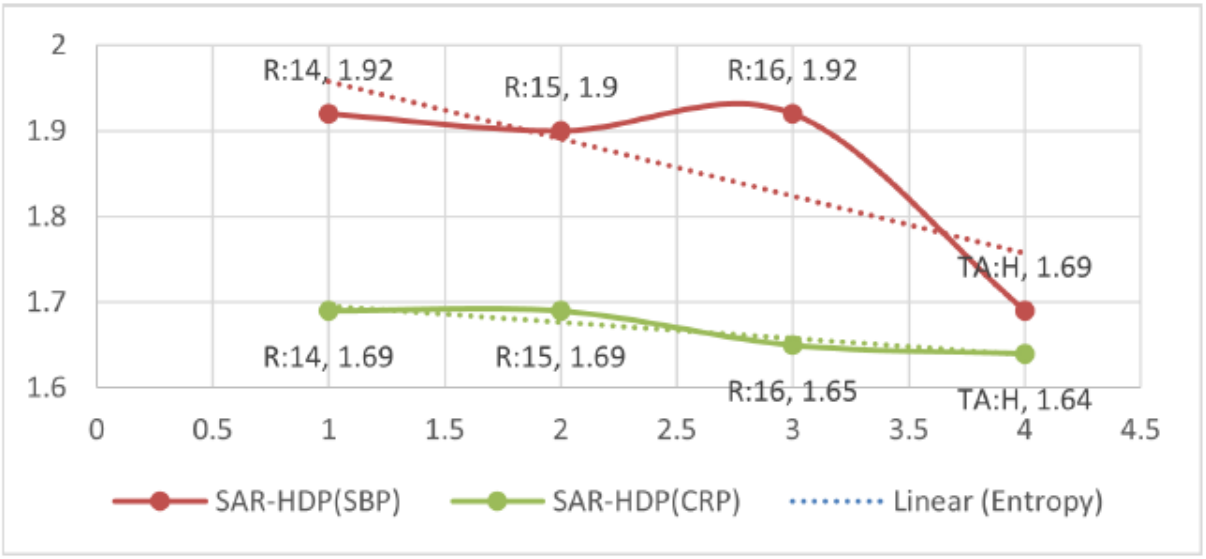

Figure : 7 Experimental results with the variation SAR-HDP settings (i.e., SBP, CRP) in terms of Entropy score

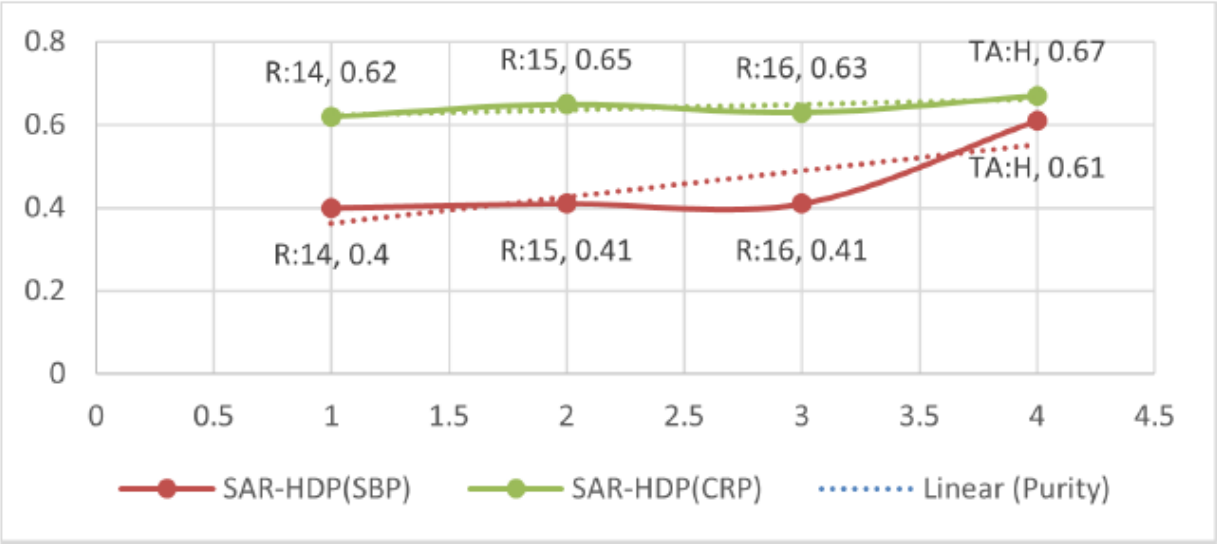

Figure : 8 Experimental results with the variation SAR-HDP settings (i.e., SBP, CRP) in terms of Purity score

6.3 Compared models

Table 7-10 examines the performance of the traditional methods (e.g., clustering methods), parametric models, and non-parametric model that are presented to be compared with the proposed SAR-HDP model in terms of aspect categorisation as follows:

6.3.1 Traditional methods

The presented traditional methods are a standard clustering algorithm such as kmeans[71], FC-Kmeans[72], and NMF[73]. None of the presented methods are possess the ability to consider the semantic regularities in the generated clusters (which are categories in this problem). However, the advanced FC-Kmeans relatively yields higher accuracy than traditional Kmeans an R:16 dataset by nearly 0.08% using RI score.

6.3.2 Parametricmodels

The parametric models in this experimental comparison are divided into standard probabilistic models that commonly built based on the frequency-based method (i.e., O-LDA [77], LDA [36], LSI [79]), and other advanced or developed models (i.e., G-LDA [78], TSLDA [46]). Thus, the proposed model produced slightly lower accuracy for aspect categorisation as opposed to the state-of-the-art 'G-LDA' ap- proach. This was attributable to the latter's status as a parametric Bayesian model, which categorised the

aspect according to its semantic information. In the case of the methods offered for aspects categorisation, topic models specifically relied on an annotated dataset due to being parametric models requiring the number of topics to be specified prior to the model training. Contrary to this, the proposed model was developed to undertake annotated and unannotated datasets alike.

6.3.3 Non-parametric models

Both of the compared non-parametric model (i.e., iLDA, O-HDP) has no mean of using semantic regularities. iLDA[70] is relatively gave higher results than O- HDP[85] model, the reason being is that iLDA relied on the Gibbs Sampling Algorithm that generate the number of topics based on a sampling algorithm, while O-HDP model built to use online versional inference as optimization algorithm that relied on (VI). Compared to our model, SAR-HDP model produced higher accuracy compared to non-parametric models.

Table 7 Comparison of RI with baselines. (the best result is in bold)

	RI				
Traditional methods	R:14	R:15	R:16	TA:H	Avg%
Kmeans	0.643	0.593	0.526	0.67	0.6
FC-Kmeans	-	-	0.5802	-	-
NMF	0.744	0.625	0.682	0.78	0.71
Parametric					
O-LDA	0.61	0.61	0.562	0.63	0.6
G-LDA	0.78	0.76	0.744	0.81	0.77
LDA	0.55	0.372	0.39	0.571	0.47
LSI	0.595	0.55	0.623	0.67	0.61
TSLDA	**0.8**	**0.83**	**0.81**	**0.87**	**0.83**
Non-parametric					
O-HDP	0.501	0.489	0.442	0.51	0.48
iLDA	0.593	0.558	0.551	0.625	0.58
SAR-HDP	0.77	0.71	0.75	0.83	0.76

Alternatively, in Table 11, the proposed model also yielded better accuracy compared to algorithm clustering (i.e. Clustering for Aspect and Feature Extrac- tion (CAFE′) [57], Non-negative Matrix Factorisation (NN-M) [75], and Mutual Reinforcement Approach (MRA) [74] implemented to perform the task on R:15 dataset in terms of RI score. CAFE′ in particular, is the second best accuracy, because it uses prior knowledge of WordNet and UMBC semantic similarity service which proves that clustering the semantic similarity between words (aspects) could have improved the aspect categories or the generated number of topics (clusters).

Table 8 Comparison of NMI with baselines. (the best result is in bold)

	NMI				
Traditionalmethods	R:14	R:15	R:16	TA:H	Avg%
Kmeans	0.093	0.066	0.037	0.208	0.1
FC-Kmeans	-	-	0.0544	-	-
NMF	0.3	0.111	0.304	0.409	0.28
Parametric					
O-LDA	0.095	0.215	0.012	0.227	0.14
G-LDA	0.278	0.421	0.256	0.402	0.34
LDA	0.121	0.121	0.115	0.144	0.12
LSI	0.028	0.3	0.1	0.341	0.19
TSLDA	**0.421**	**0.482**	**0.486**	**0.656**	**0.51**
Non-parametric					
O-HDP	0.089	0.011	0.012	0.394	0.12
iLDA	0.224	0.211	0.22	0.482	0.28
SAR-HDP	0.4	0.31	0.34	0.52	0.39

Table 9 Comparison of Entropy with baselines. (the best result is in bold)

	Entropy				
Traditional methods	R:14	R:15	R:16	TA:H	Avg%
Kmeans	1.32	2.013	1.741	1.021	1.52
FC-Kmeans	-	-	1.6592	-	-
NMF	1.32	1.136	1.119	0.99	1.14
Parametric					
O-LDA	1.332	1.452	1.4	1.006	1.3
G-LDA	1.21	1.001	1.181	0.989	1.1
LDA	1.768	1.784	1.725	1.2	1.62
LSI	2.125	1.458	1.359	1.032	1.49
TSLDA	**1.087**	**0.976**	**1**	**0.957**	**1**
Non-parametric					
O-HDP	2.015	2.015	2.096	2.096	2.05
iLDA	1.905	1.905	1.963	1.632	1.85
SAR-HDP	1.69	1.69	1.65	1.64	1.66

Table 10 Comparison of Purity with baselines. (the best result is in bold)

	Purity				
Traditional methods	R:14	R:15	R:16	TA:H	Avg%
Kmeans	0.467	0.489	0.432	0.511	0.47
FC-Kmeans	-	-	0.617582	-	-
NMF	0.625	0.522	0.442	0.622	0.55
Parametric					
O-LDA	0.45	0.43	0.436	0.421	0.43
G-LDA	0.595	0.591	0.594	0.532	0.58
LDA	0.669	0.492	0.472	0.604	0.56
LSI	0.478	0.364	0.511	0.421	0.44
TSLDA	**0.679**	**0.688**	**0.587**	**0.712**	**0.67**
Non-parametric					
O-HDP	0.492	0.495	0.559	0.61	0.53
iLDA	0.56	0.55	0.573	0.63	0.57
SAR-HDP	0.62	0.65	0.63	0.67	0.64

Table 11 Comparison of RI score with baselines on R:15 dataset.

Methods	RI
MRA	0.61
CAFÉ	0.68
NN-M	0.62
SAR-HDP	**0.71**

Table 12 presented a comparison of a supervised method FFDML[76] compared to AP[55] and L-EM[55] that are replicated in [55]. These methods produced lower accuracy in terms of RI, NMI, and Purity on the R:16 dataset in comparison to SAR-HDP model.

Table 12 Comparison of RI, NMI, Entropy, and Purity scores with baselines on R:16 dataset.

Methods	RI	NMI	Entropy	Purity
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
SAR-HDP	**0.75**	**0.34**	1.65	**0.63**

However, the proposed model outperformed in the same category across dif- ferent aspect categorisation methods pertaining to extracted aspect coverage in each category. The information included in Table 13 summarises the performance shown by the current methods according to their RI score. Apart from the Topic- seeds LDA (TSLDA) and Sentic LDA [39] topic models, performance displayed by the proposed model exceeded all previously proposed methods, DF-LDA [80], SAS [81], SJASM [38], sLDA [82], JST [83], ASUM [84], LARA [65] for aspect categorisation in the TA:H dataset.

Table 13 Comparison of RI, and Entropy scores with baselines on TA:H dataset.

Methods	RI	Entropy
DF-LDA	0.75	1.75
SAS	0.77	1.45
SenticLDA	**0.86**	**0.95**
sLDA	0.72	-
JST	0.7	-
ASUM	0.69	-
LARA	0.63	-
SAR-HDP	0.83	1.64

7. Limitations and future work

The proposed SAR-HDP model suffers from some limitations that make it not sufficient to deal with all the aspect-terms, because the developed SAR model don't consider having the aspect terms that are not articulated SAR rules. Future direction is to consider the Dependency-parser and the rule-based combined.

8. Conclusion

In the breadth of massively generated sentimental reviews across all public plat- forms about any particular products or services, the need for a non-parametric model capable of extracting and categorising aspect terms automatically with- out necessitating a class label cannot be denied. To this end, supervised machine learning and parametric models are commonly associated with the difficulties and challenges of automatically grasping the semantic associations between aspects and their expressed sentiments, whereby these aspects are then placed into categories accordingly.

To address the identification of aspect(s) and recognising their respective as- pect topic(s) (i.e., category), this work proposed a non-parametric model named SAR-HDP. It extended beyond considering the semantic association between the aspect(s) and its expressed sentiment(s) to also allocate them into aspect topic(s) accordingly. The model could be differentiated from conventional methods due to its lack of consideration for document/sentence review drawn from a single topic; instead, it would allocate varying aspect(s) within a single sentence to sin- gular or multiple topic(s). Here, the SAR-HDP model was built on the premise of syntactical relation for the identification of aspect-sentiment pair, which also dismissed irrelevant aspects by assessing the semantic assertions between them. In particular, the SAR developed functioned in tandem with the HDP model for the purpose of aspect identification and allocation into categories. Therefore, draws made for the model categories were reliant upon the MGD model, which consid- ered semantic regularities by using the space of aspects and categories offered by embeddings. Last but not least, the hyperparameter values were then discovered using a sampling algorithm (Gamma distribution).

Funding

Not applicable.

Conflict of interest

The authors declare that they have no conflict of interest.

Code availability

Not applicable.

Availability of data and material

Datasets are available as follows:
TripAdvisor: Hotel (TA:H)
SemEval-2014 Restaurant
SemEval-2015 Restaurant
SemEval-2016 Restaurant

References

1. Liu Y, Bi JW, Fan ZP (2017) Ranking products through online reviews: A method based on sentiment analysis technique and intuitionistic fuzzy set theory. Information Fusion 36:149–161. https://doi.org/10.1016/j.inffus.2016.11.012
2. Cambria E (2016) Affective Computing and Sentiment Analysis. IEEE Intelligent Systems 31:102–107. https://doi.org/10.1109/MIS.2016.31
3. Rana TA, Cheah Y-N (2016) Aspect extraction in sentiment analysis: comparative analysis and survey. Artificial Intelligence Review 46:459–483
4. Hu M, Liu B (2004) Mining and summarizing customer reviews. Proceedings of the 2004 ACM SIGKDD international conference on Knowledge discovery and data mining (KDD) 168. https://doi.org/10.1145/1014052.1014073
5. Popescu AM, Etzioni O (2005) Extracting product features and opinions from reviews. In: HLT/EMNLP 2005 - Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing, Proceedings of the Conference. Springer London, pp 339–346
6. Moghaddam S, Ester M (2010) Opinion digger: an unsupervised opinion miner from unstructured product reviews. In: Proceedings of the 19th ACM international conference on Information and knowledge management - CIKM '10. ACM Press, New York, New York, USA, p 1825
7. Htay SS, Lynn KT (2013) Extracting Product Features and Opinion Words Using Pattern Knowledge in Customer Reviews. The Scientific World Journal 2013:1–5

8. Hai Z, Chang K, Kim J-J, Yang CC (2014) Identifying Features in Opinion Mining via Intrinsic and Extrinsic Domain Relevance. IEEE Transactions on Knowledge and Data Engineering 26:623–634
9. Poria S, Cambria E, Ku L-W, et al (2015) A Rule-Based Approach to Aspect Extraction from Product Reviews. In: second workshop on natural language processing for social media (SocialNLP). pp 28–37
10. Rana TA, Cheah YN (2015) Hybrid rule-based approach for aspect extraction and categorization from customer reviews. In: 2015 9th International Conference on IT in Asia: Transforming Big Data into Knowledge, CITA 2015 - Proceedings. IEEE, pp 1–5
11. Wu Y, Zhang Q, Huang X, Wu L (2009) Phrase dependency parsing for opinion mining. Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing Volume 3 - EMNLP '09 3:1533. https://doi.org/10.3115/1699648.1699700
12. Wei CP, Chen YM, Yang CSCC, Yang CSCC (2010) Understanding what concerns consumers: A semantic approach to product feature extraction from consumer reviews. Information Systems and e-Business Management 8:149–167. https://doi.org/10.1007/s10257-009-0113-9
13. Samha, A. K., Li, Y., Zhang, J. (2014). Aspect-based opinion extraction from customer reviews. arXiv preprint arXiv:1404.1982.
14. Liu, K., Xu, L., Zhao, J. (2013). Syntactic patterns versus word alignment: Extracting opinion targets from online reviews. In Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers) (pp. 1754-1763).
15. Yan Z, Xing M, Zhang D, Ma B (2015) EXPRS: An extended pagerank method for product feature extraction from online consumer reviews. Information and Management 52:850–858. https://doi.org/10.1016/j.im.2015.02.002
16. Choi, Y, Cardie, C. (2010). Hierarchical sequential learning for extracting opinions and their attributes. In Proceedings of the ACL 2010 conference short papers (pp. 269-274).
17. Jin W, Ho HH (2009) A novel lexicalized HMM-based learning framework for web opinion mining. In: ACM International Conference Proceeding Series. ACM Press, New York, New York, USA, pp 1–8
18. Chen L, Qi L, Wang F (2012) Comparison of feature-level learning methods for mining online consumer reviews. Expert Systems with Applications 39:9588–9601. https://doi.org/10.1016/j.eswa.2012.02.158
19. Venugopalan M, Gupta D, Bhatia V (2021) A Supervised Approach to Aspect Term Extraction Using Minimal Robust Features for Sentiment Analysis. In: Advances in Intelligent Systems and Computing. Springer, Singapore, pp 237–251
20. Hu, M., Liu, B. (2004). Mining opinion features in customer reviews. In AAAI (Vol. 4, No. 4, pp. 755-760).
21. Li S, Zhou L, Li Y (2015) Improving aspect extraction by augmenting a frequency-based method with web-based similarity measures. Information Processing and Management 51:58–67. https://doi.org/10.1016/j.ipm.2014.08.005
22. Rana TA, Cheah Y (2018) Improving Aspect Extraction Using Aspect Frequency and Semantic Similarity-Based Approach for Aspect-Based Sentiment Analysis. 566:
23. Rana TA, Cheah YN (2019) Sequential patterns rule-based approach for opinion target extraction from customer reviews. Journal of Information Science 45:643–655. https://doi.org/10.1177/0165551518808195
24. Chauhan GS, Meena YK (2020) DomSent: Domain-Specific Aspect Term Extraction in Aspect-Based Sentiment Analysis. In: Smart Innovation, Systems and Technologies. Springer, pp 103–109
25. Qiu G, Liu B, Bu J, Chen C (2011) Opinion Word Expansion and Target Extraction through Double Propagation. Computational Linguistics, MIT press 37:9–27

26. Liu Q, Gao Z, Liu B, Zhang Y (2015) Automated rule selection for aspect extraction in opinion mining. IJCAI International Joint Conference on Artificial Intelligence 2015-Janua:1291–1297
27. Ravi Kumar V, Raghuveer K (2013) Dependency driven semantic approach to product features extraction and summarization using customer reviews. In: Advances in Intelligent Systems and Computing. Springer Verlag, pp 225–238
28. Liu Q, Gao Z, Liu B, Zhang Y (2016) Automated rule selection for opinion target extraction. Knowledge-Based Systems 104:74–88. https://doi.org/10.1016/j.knosys.2016.04.010
29. Kang Y, Zhou L (2017) RubE: Rule-based methods for extracting product features from online consumer reviews. Information and Management 54:166–176. https://doi.org/10.1016/j.im.2016.05.007
30. Maharani W, Widyantoro DH, Khodra ML (2015) SAE: Syntactic-based aspect and opinion extraction from product reviews. In: ICAICTA 2015 - 2015 International Conference on Advanced Informatics: Concepts, Theory and Applications. IEEE, pp 1–6
31. Asghar MZ, Khan A, Zahra SR, et al (2019) Aspect-based opinion mining framework using heuristic patterns. Cluster Computing 22:7181–7199. https://doi.org/10.1007/s10586-017-1096-9
32. Rana TA, Cheah Y-N (2017) A two-fold rule-based model for aspect extraction. Expert Systems with Applications 89:273–285
33. Rana TA, Cheah YN, Rana T (2020) Multi-level knowledge-based approach for implicit aspect identification. Applied Intelligence 50:4616–4630. https://doi.org/10.1007/s10489-020-01817-x
34. Jelodar H, Wang Y, Yuan C, et al (2019) Latent Dirichlet allocation (LDA) and topic modeling: models, applications, a survey. Multimedia Tools and Applications 78:15169–15211. https://doi.org/10.1007/s11042-018-6894-4
35. Rana TA, Cheah YN, Letchmunan S (2016) Topic modeling in sentiment analysis: A systematic review. Journal of ICT Research and Applications 10:76–93. https://doi.org/10.5614/itbj.ict.res.appl.2016.10.1.6
36. Blei DM, Ng AY, Jordan MI (2003) Latent Dirichlet allocation. Journal of Machine Learning Research 3:993–1022
37. Bagheri A, Saraee M, De Jong F (2014) ADM-LDA: An aspect detection model based on topic modelling using the structure of review sentences. Journal of Information Science 40:621–636
38. Hai Z, Cong G, Chang K, et al (2017) Analyzing sentiments in one go: A supervised joint topic modeling approach. IEEE Transactions on Knowledge and Data Engineering 29:1172–1185. https://doi.org/10.1109/TKDE.2017.2669027
39. Poria S, Chaturvedi I, Cambria E, Bisio F (2016) Sentic LDA: Improving on LDA with semantic similarity for aspect-based sentiment analysis. In: International Joint Conference on Neural Networks (IJCNN). IEEE, pp 4465–4473
40. Shams M, Baraani-Dastjerdi A (2017) Enriched LDA (ELDA): Combination of latent Dirichlet allocation with word co-occurrence analysis for aspect extraction. Expert Systems With Applications, Elsevier 136–146
41. Chen Z, Mukherjee A, Liu B, et al (2013) Discovering coherent topics using general knowledge. In: international Conference on information & knowledge management, ACM. pp 209–218
42. García-Pablos A, Cuadros M, Rigau G (2018) W2VLDA: Almost unsupervised system for Aspect Based Sentiment Analysis. Expert Systems with Applications 91:127–137. https://doi.org/10.1016/j.eswa.2017.08.049
43. Yuan B, Wu G (2017) A Hybrid HDP-ME-LDA Model for Sentiment Analysis. In: Proceedings of the 2017 2nd International Conference on Automation, Mechanical Control and Computational Engineering (AMCCE 2017). pp 659–663

44. Brody S, Elhadad N (2010) An Unsupervised Aspect-Sentiment Model for Online Reviews. In: HLT '10 Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics. pp 804–812
45. Wan C, Peng Y, Xiao K, et al (2020) An association-constrained LDA model for joint extraction of product aspects and opinions. Information Sciences 519:243–259
46. Al-Janabi OM, Malim NHAH, Cheah YN (2020) Aspect Categorization Using Domain-Trained Word Embedding and Topic Modelling. In: Lecture Notes in Electrical Engineering. Springer, Singapore, pp 191–198
47. Ding W, Song X, Guo L, et al (2013) A novel hybrid HDP-LDA model for sentiment analysis. In: Proceedings - 2013 IEEE/WIC/ACM International Conference on Web Intelligence, WI 2013. pp 329–336
48. Al-Janabi OM, Malim NHAH, Cheah YN (2018) Analogy-based Common-Sense Knowledge for Opinion-Target Identification and Aggregation. In: 2018 2nd International Conference on Imaging, Signal Processing and Communication, ICISPC 2018. Institute of Electrical and Electronics Engineers Inc., pp 1–5
49. Zheng X, Lin Z, Wang X, et al (2014) Incorporating appraisal expression patterns into topic modeling for aspect and sentiment word identification. Knowledge-Based Systems, Elsevier 29–47
50. Park S-M, Lee SJ, On B-W (2020) Topic Word Embedding-Based Methods for Automatically Extracting Main Aspects from Product Reviews. Applied Sciences 10:3831. https://doi.org/10.3390/app10113831
51. Yang M, Hsu WH (2013) HDPsent: Incorporation of Latent Dirichlet Allocation for Aspect_Level Sentiment into Hierarchical Dirichlet Process-Based Topic Models. Journal of Chemical Information and Modeling 53:1689–1699.https://doi.org/10.1017/CBO9781107415324.004
52. Zhang W, Xu H, Wan W (2012) Weakness Finder: Find product weakness from Chinese reviews by using aspects based sentiment analysis. Expert Systems with Applications, Elsevier 10283–10291
53. Kiritchenko S, Zhu X, Cherry C, Mohammad S (2014) NRC-Canada-2014: Detecting Aspects and Sentiment in Customer Reviews. In: Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014). pp 437–442
54. Xu Q, Zhu L, Dai T, et al (2020) Non-negative matrix factorization for implicit aspect identification. Journal of Ambient Intelligence and Humanized Computing 11:2683–2699. https://doi.org/10.1007/s12652-019-01328-9
55. Zhai Z, Liu B, Xu H, Jia P (2011) Clustering product features for opinion mining. Proceedings of the 4th ACM International Conference on Web Search and Data Mining, WSDM 2011 347–354. https://doi.org/10.1145/1935826.1935884
56. Pessutto LRC, Vargas DS, Moreira VP (2020) Multilingual aspect clustering for sentiment analysis. Knowledge-Based Systems 192:105339. https://doi.org/10.1016/j.knosys.2019.105339
57. 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 of the Association for Computational Linguistics: Human Language Technologies, NAACL HLT 2016 - Proceedings of the Conference. pp 789–799
58. Mikolov T, Sutskever I, Chen K, et al (2013) Distributed representations of words and phrases and their compositionality. Advances in Neural Information Processing Systems 1–9
59. Sidorov G, Gelbukh A, Gómez-Adorno H, Pinto D (2014) Soft similarity and soft cosine measure: Similarity of features in vector space model. Computacion y Sistemas 18:491–504. https://doi.org/10.13053/CyS-18-3-2043
60. Li C, Rana S, Phung D, Venkatesh S (2016) Hierarchical Bayesian nonparametric models for knowledge discovery from electronic medical records. Knowledge-Based Systems 99:168–182. https://doi.org/10.1016/j.knosys.2016.02.005
61. Aldous DJ (1985) Exchangeability and related topics. Springer, Berlin, Heidelberg, pp 1–198

62. Rand WM (1971) Objective criteria for the evaluation of clustering methods. Journal of the American Statistical Association 66:846–850. https://doi.org/10.1080/01621459.1971.10482356
63. Shannon C (1948) A mathematical theory of communication — Nokia Bell Labs Journals & Magazine — IEEE Xplore. In: The Bell System Technical Journal. 10.1002/j.1538-7305.1948.tb01338.x. Accessed 27 Apr 2021
64. Strehl A, Ghosh J (2003) Cluster ensembles - A knowledge reuse framework for combining multiple partitions. In: Journal of Machine Learning Research. pp 583–617
65. Wang H, Lu Y, Zhai C (2010) Latent aspect rating analysis on review text data: A rating regression approach. In: Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '09. ACM Press, New York, USA, pp 783–792
66. Giordano, R., Liu, R., Jordan, M. I., & Broderick, T. (2022). Evaluating sensitivity to the stick-breaking prior in Bayesian nonparametrics. Bayesian Analysis, 1(1), 1-34.
67. Batmanghelich NK, Saeedi A, Narasimhan KR, Gershman SJ (2016) Nonparametric spherical topic modeling with word embeddings. In: 54th Annual Meeting of the Association for Computational Linguistics, ACL 2016 - Short Papers. Association for Computational Linguistics (ACL), pp 537–542
68. Jameel S, Schockaert S (2020) Word and document embedding with VMF-mixture priors on context word vectors. In: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. pp 3319–3328
69. Reisinger J, Waters A, Silverthorn B, Mooney RJ (2010) Spherical topic models. In: ICML 2010 - Proceedings, 27th International Conference on Machine Learning. pp 903–910
70. Heinrich G (2011) "Infinite LDA" – Implementing the HDP with minimum code complexity. ArbylonNet 1–20
71. Arthur D, Vassilvitskii S (2007) K-means++: The advantages of careful seeding. In: Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms
72. Xiong S, Ji D (2016) Exploiting flexible-constrained K-means clustering with word embedding for aspect-phrase grouping. Information Sciences 367–368:689–699. https://doi.org/10.1016/j.ins.2016.07.002
73. Hoyer PO (2004) Non-negative matrix factorization with sparseness constraints. Journal of Machine Learning Research 5:1457–1469
74. Su Q, Xu X, Guo H, et al (2008) Hidden sentiment association in Chinese web opinion mining. In: Proceeding of the 17th International Conference on World Wide Web. pp 959–968
75. Hai Z, Chang K, Kim JJ (2011) Implicit feature identification via co-occurrence association rule mining. In: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). pp 393–404
76. Xiong S, Cheng M, Batra V, et al (2020) Aspect terms grouping via fusing concepts and context information. Information Fusion 64:12–19. https://doi.org/10.1016/j.inffus.2020.06.007
77. Hoffman MD, Blei DM, Bach F (2010) Online learning for Latent Dirichlet Allocation. In: Advances in Neural Information Processing Systems 23 (NIPS). pp 856–864
78. Das R, Zaheer M, Dyer C (2015) Gaussian LDA for topic models with word embeddings. In: ACL-IJCNLP 2015 - 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing of the Asian Federation of Natural Language Processing, Proceedings of the Conference. pp 795–804
79. Dumais ST (2004) Latent Semantic Analysis. Annual Review of Information Science and Technology 188–230
80. Andrzejewski D, Zhu X, Craven M (2009) Incorporating domain knowledge into topic modeling via Dirichlet Forest priors. In: Proceedings of the 26th Annual International Conference on Machine Learning - (ICML). ACM, New York, New York, USA, pp 1–8
81. Mukherjee A, Liu B (2012) Aspect extraction through semi-supervised modeling. In: Proceedings of 50th Annual Meeting of the Association for Computational Linguistics, (ACL). Association for Computational Linguistics, pp 339–348

82. Blei DM, McAuliffe JD (2009) Supervised topic models. Advances in Neural Information Processing Systems 121–128
83. Lin C, He Y (2009) Joint sentiment/topic model for sentiment analysis. In: Proceeding of the 18th ACM conference on Information and knowledge management - (CIKM). ACM Press, p 375
84. Jo Y, Oh A (2011) Aspect and sentiment unification model for online review analysis. In: Proceedings of the 4th ACM International Conference on Web Search and Data Mining, (WSDM). ACM Press, pp 815–824
85. Wang C, Paisley J, Blei DM (2011) Online Variational Inference for the Hierarchical Dirichlet Process. In: Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics, PMLR. pp 752–760