

# Integrating word status for joint detection of sentiment and aspect in reviews

Journal of Information Science  
2019, Vol. 45(6) 736–755  
© The Author(s) 2018



Article reuse guidelines:

[sagepub.com/journals-permissions](http://sagepub.com/journals-permissions)

DOI: 10.1177/0165551518811458

[journals.sagepub.com/home/jis](http://journals.sagepub.com/home/jis)



**Ayoub Bagheri**

Department of Methodology and Statistics, Faculty of Social Sciences, Utrecht University, The Netherlands;  
Division of Heart and Lungs, Department of Cardiology, University Medical Center Utrecht, The Netherlands

## Abstract

A crucial task in sentiment analysis is aspect detection: the step of selecting the aspects on which opinions are expressed. This step anticipates the step of determining whether the opinions on aspects are positive or negative. This article proposes a novel probabilistic generative topic model for aspect-based sentiment analysis which is able to discover the latent structure of a large collection of review documents. The proposed joint sentiment-aspect detection model (SAM) is a generative topic model that incorporates the structure of review sentences for detecting aspects and sentiments simultaneously. The intuitions behind the SAM are that from generating documents by latent single- and multi-word topics, modelling the word distribution for each topic and learning of the prior distribution over topics in sentences of documents. SAM introduces word status so that the model can decide when to sample from a bigram distribution or a unigram distribution and integrates all these components into one combined model for aspect-based sentiment analysis. We evaluate SAM both qualitatively and quantitatively to show that the model is indeed able to perform the task effectively and improves significantly over standard joint sentiment-aspect models. The proposed model can easily be transformed between domains or languages and can detect the polarity of text data at various levels. However, for the quantitative analysis, we mainly focus on presenting the results for the document-level sentiment classification.

## Keywords

Aspect-based sentiment analysis; joint sentiment aspect; latent Dirichlet allocation; online reviews; sentiment analysis; topic modelling

## 1. Introduction

With the growing popularity of the Web, the number of reviews containing opinions, feedbacks, comments, appraisals and so on is increasing at a high speed [1]. Such reviews provide valuable information which can help people to make their decisions on whether or not to purchase a product or service [1,2]. For example, customers' reviews are essential to other potential customers, retailers and product manufacturers in their efforts to understand the general opinions of customers and help them to make better decisions [3]; assessments on politics can help politicians to interpret the situation and plan their strategy, for example, governments want to know what people who have voted for them think about their activities; and restaurant reviews can assist customers to pick out a place based on their needs [4–6]. As the number of reviews is expanding, it becomes very hard for users to obtain a comprehensive view of others' opinions about various aspects of objects, products or services through a manual analysis [1,2,7–10]. A proper analysis of the opinions expressed and a summarisation can enable potential users to check previous positive and negative opinions about specific features or aspects of objects [1,10]. This functionality is commonly referred to as sentiment analysis.

In the past few years, sentiment analysis has become a rapidly developing field of study, aiming at models and approaches for the automated analysis of people's opinions expressed in reviews [1,11–18]. Sentiment analysis may involve techniques from a number of fields including text mining, natural language processing and computational linguistics [1].

---

## Corresponding author:

Ayoub Bagheri, Department of Methodology and Statistics, Faculty of Social Sciences, Utrecht University, Utrecht 3508 TC, The Netherlands.  
Email: [a.bagheri@uu.nl](mailto:a.bagheri@uu.nl); [a.bagheri-2@umcutrecht.nl](mailto:a.bagheri-2@umcutrecht.nl)

A crucial subtask in sentiment analysis is aspect detection: isolation of the aspects on which opinions are expressed [3,5,8,14,19]. This step anticipates the step of determining whether the opinions on aspects are positive or negative [1,4,6,7] and is related to information extraction and consists of extracting structured representations of aspect-oriented opinions from review documents [12,13,18]. In the context of reviews, aspects can be considered to coincide with the topics on which opinions are expressed and sentiment-aspect classification determines whether the opinions on the detected aspects (or topics) are positive, negative or neutral [14,19]. In the remainder of this article, we will use the notion of aspect and topic interchangeably and the focus will be on detecting aspects and determining the orientation of the sentiment expressed for the aspects in a review. In terms of the approach adopted, there are three types of approaches for aspect detection and sentiment-aspect classification tasks [1]: (1) the supervised learning approach, (2) the frequency- and relation-based approach and (3) the topic modelling-based approach.

Up till now, many supervised learning algorithms such as support vector machines, naïve Bayes, neural networks and decision trees have been proposed for the sentiment analysis tasks [4,20–22]. Supervised algorithms are dependent on a set of pre-labelled training data. Although the supervised approaches can achieve reasonable effectiveness, building sufficient labelled data is often expensive and needs much human labour [3]. In addition, a model trained on labelled data in one domain often performs poorly in another domain [1]. Frequency- and relation-based approaches can analyse the opinions in review documents in an unsupervised or semi-supervised way [5,14]; hence, these approaches overcome the need for labelled data that were inherent to the supervised learning. The major shortcoming of the frequency- and relation-based approaches is that they require manual tuning of various parameters, which makes them hard to transfer to another domain [1,14]. In contrast, topic modelling-based methods are unsupervised approaches which benefit from the relation between words and the topics [23–28]. The basic idea in topic modelling is that documents are represented as random mixtures over latent topics, where each topic is characterised by a distribution over words [29–32]. One main advantage of topic modelling-based approaches is that these methods view the text as a mixture of global aspects (or topics) and then analyse the sentiment in the more detailed topic or domain level [27,33].

Due to wide range of products and services being reviewed on the Internet, supervised, domain-specific or language-dependent models are not often practical [5,14,27], and it is desirable to develop models that work with unlabeled data [23]. Therefore, the framework for aspect-based sentiment analysis must be robust and easily transferable between domains or languages. Unsupervised topic modelling using approaches such as latent Dirichlet allocation (LDA) for the aspect-based sentiment analysis has encountered considerable popularity as a way to model latent aspects and sentiments in textual data [23]. Although topic models seem to benefit from the correlation between words and topics, the assumption inherent to these models that texts are a bag of words and that the order of words in a sentence can be ignored is questionable. In this article, we present a novel unsupervised model based on topic modelling which addresses the core tasks necessary to detect aspects and sentiments from review sentences for a document-level sentiment analysis system. The proposed model is expected to yield better results by the fact that it relaxes the bag-of-words assumption by considering word status. Improvements are predicted for the inferring of latent aspects regarding structure of sentences in a document, the ability to extract multi-word aspects and the power to predict the topic of previously unobserved words in documents.

One of the main issues in aspect-based sentiment analysis is that in review sentences, the sentiment polarities are dependent on aspects, topics or domains [27,28]. For example, the opinion (sentiment) word ‘*small*’ in the review sentence ‘*after using iPod, I found the size is small and perfect for carrying in my pocket*’ has a positive orientation on the aspect ‘*size*’ and could have a negative orientation in the sentence ‘*I don’t like the screen, it is too small!*’. Hence, sentiment words could have different meanings or orientations in different domains. To overcome this issue, the model proposed in this article tries to extract both aspect and sentiment simultaneously. The proposed joint sentiment-aspect detection model (SAM) approach is based on the concept of a generative topic model that incorporates word status and structure of review sentences for detecting aspects and sentiments simultaneously. SAM differs from existing techniques in that (1) SAM requires no labelled training data; (2) SAM can detect aspect and sentiment simultaneously; (3) SAM assigns all words in a sentence to one aspect; and (4) SAM utilises word status and the structure of reviews, that is, the order of words, syntax information, word co-occurrences and frequencies of words, to extract single- and multi-word aspects. SAM tries to extract latent aspects and sentiments from reviews by making use of information of documents (e.g. word co-occurrences and frequencies of words) as well as the order of words in each document. As it is an unsupervised approach, SAM can easily be transformed between domains or languages. This model can detect polarity of text data at various levels – that is, document, sentence or aspect level – however for the qualitative analysis, we present several generated aspects and sentiments and for the quantitative analysis, we mainly focus on presenting the results for the document-level sentiment classification.

The remainder of this article is organised as follows. The detailed discussions of existing work are given in section 2. Section 3 proceeds by reviewing the formalism of LDA. Section 4 describes the proposed SAM model, including the overall architecture and specific design aspects. Subsequently, we describe an empirical evaluation and discuss the major experimental results in section 5. Finally, we conclude with a summary and some future research directions in section 6.

## 2. Related work

The phrase ‘sentiment analysis’ perhaps first appeared in the study of Yi et al. [34]. However, studies of sentiment analysis or opinion mining appeared earlier [20,35]. In this article, we use the terms sentiment and opinion interchangeably. The term sentiment can denote opinion, sentiment, evaluation, appraisal, attitude and emotion [1]. Aspect-based sentiment analysis mainly focuses on detecting aspects and opinions which express or imply positive or negative sentiments [1,27].

Various approaches have been proposed for aspect-based sentiment analysis of review texts. Previous work is based on, for example, association rule mining [2], double propagation [10], unsupervised aspect detection [5,14] and supervised learning methods [4]. These approaches have limitation that they do not group semantically related aspect expressions together [1,25,36]. However, recent approaches using deep neural networks [22,37] showed significant performance improvement over the state-of-the-art supervised methods on a range of sentiment analysis tasks. Soujanya et al. [22] deployed a seven-layer deep convolutional neural network for aspect-based sentiment analysis to tag each word in opinionated sentences as either aspect or non-aspect word. In addition, supervised methods are often not practical due to the fact that creating sufficient volumes of labelled data is often expensive and needs much human labour [3]. Unsupervised topic modelling approaches for identifying aspect words have been shown to be more effective [23,27,36,38]. They involve topic models and consist of a suite of probabilistic algorithms whose aim is to extract latent structure from large collection of documents [29,30]. These models all share the idea that documents are mixtures of topics and that each topic is a distribution over words [32].

Current topic modelling approaches are computationally efficient and seem to capture correlations between words and topics, but they have one main limitation [39–43]. This limitation is that they assume that words are generated independently of each other. This is known as the bag-of-words assumption [39]. In other words, topic models only extract unigrams for topics in a corpus. The bag-of-words assumption for current topic modelling approaches ignores the order of words [40]. This is an unrealistic simplification. In the past few years, several studies have been published that use topic models for the problem of aspect-based sentiment analysis to overcome this limitation [23,39,40,42,44]. Wallach [39] developed a bigram topic model based on the hierarchical Dirichlet language model, using a hierarchical Bayesian model that integrates bigram- and topic-based techniques to document modelling. Wallach’s model does not consider unigram aspects and always generates bigrams. Griffiths et al. [44] proposed a model named the LDA collocation model. It introduces a new set of parameters to decide whether to generate a unigram or a bigram aspect, whereas their model does not always generate a reasonable topic for a word or phrase. Wang et al. [42] improved the LDA collocation model to make it possible to decide whether to form an  $n$ -gram for the successive words depending on their nearby context and co-occurrences. All these models assume that the words in a document are generated by a latent topic assignment with regard to  $n$ -previous words in the document. Gruber et al. [40] used the same line of research as well as assuming Markovian relations between latent aspects. They proposed a hidden topic Markov model in which all words of a sentence have the same topic and successive sentences are more likely to have the same topics. Although using order of words in hidden topic Markov model does extract topics, it does not work well for extracting  $n$ -gram aspects. In addition, all these models only try to detect aspects from textual data, while another important task in aspect-based sentiment analysis is the discovery of opinions and sentiments for the detected aspects [1].

Motivated by these observations, we follow this promising line of research. This article is based on the assumption that a topic model considering unigrams and phrases jointly with extracting sentiments is more realistic and more useful in applications. Several integrated models of topics and sentiments have been proposed [26–28,33,45–47]. These models extend from the basic topic models’ idea that an opinionated document is a mixture of topics and a topic is a distribution over words. Topic-sentiment model (TSM) [26] is one of the first proposed approaches which jointly models the mixture of topics and sentiments for documents. There are several differences between TSM and our model. TSM is based on probabilistic latent semantic indexing (PLSI) [30] and has shortcomings on inferencing for new documents and overfitting the data. Our LDA-based model overcomes this. Also, TSM represents sentiment through a language model separately from topics, while SAM considers a topic-sentiment pair as a single unit for the language model. Titov and McDonald [19] proposed a multi-aspect sentiment model (MAS) which is claimed to extract topics that are representative of ratable aspects from user reviews and can obtain sentiment summary for each aspect. Our model differs from MAS in several ways: MAS is a supervised sentiment-topic model as it requires that every aspect is rated in training

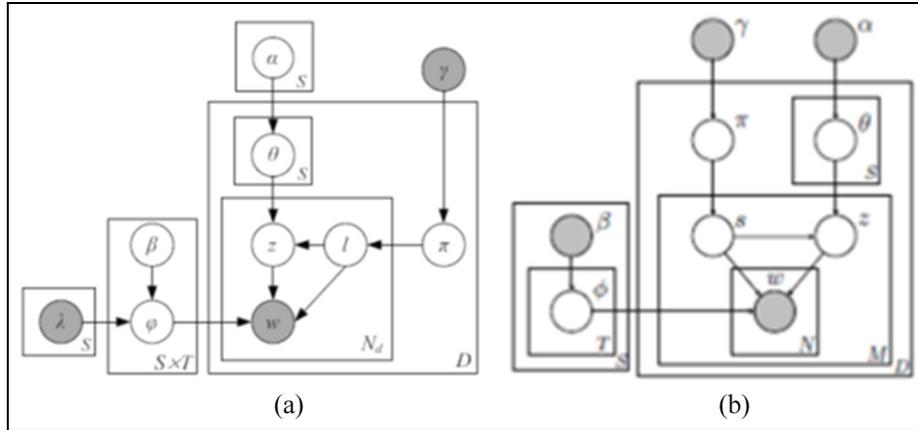


Figure 1. (a) JST model and (b) ASUM model.

documents, while SAM is a weakly supervised model with minimum seed information being incorporated and does not use any labelled training data. MAS tries to extract sentiment summaries for each aspect, while our model can present the outcome in aspect, sentence or document level. Both the joint sentiment-topic model (JST) [27] and the aspect-sentiment unification model (ASUM) [28] shown in Figure 1 are highly similar to the SAM model. In these models, sentiments are unified with topics in a single language model. Both JST and ASUM are based on LDA and are fully unsupervised approaches including only minimum prior information for sentiment-aspect classification. JST views topics and sentiments from a document level, but ASUM limits the words in sentence level to come from the same language model. ASUM relaxes the assumption that the subsequent words in a document or a sentence have different aspects, but it does not utilise the structure of sentences in terms of word order, nor does it focus on the extraction of multi-word aspects from text data. Our model, SAM, relaxes the bag-of-words assumption for the extraction of unigram and  $n$ -gram aspect phrases by assuming that the aspects of words in a sentence form a Markov chain and subsequent words are more likely to have the same aspect. Also, the proposed model takes a small set of 20 general affective and evaluative words and detects aspect-specific sentiment words jointly with discovering aspects. Therefore, in this article, we propose a new topic modelling approach that can automatically extract aspects and sentiments by structure of review sentences, that is, using word frequencies, order of words and syntax of sentences. SAM is designed for aspect-based sentiment analysis to be as unsupervised as possible, to make it transferable through different domains, as well as across languages.

### 3. LDA

Topic models are based on the assumption that documents are mixture of topics, where a topic is a probability distribution over words. A topic model is a generative model for documents where it specifies a probabilistic procedure based on probabilistic sampling rules that describe how words in documents might be generated on the basis of random variables [32]. LDA is one of the most popular topic models where its probabilistic procedure ties parameters of documents via a hierarchical generative model [29].

Figure 2 shows the graphical model of the LDA [29]. In this graphical notation, nodes are random variables and edges indicate conditional dependencies between variables. Shaded and unshaded variables indicate observed and latent (i.e. unobserved and hidden) variables, respectively, while plates refer to repetitions of sampling steps with the variable in the lower right corner referring to the number of samples [32].

Given a corpus with a collection of  $D$  documents, LDA assumes that each word  $w$  is associated with a latent topic  $z$ . Each of these topics  $j \in \{1, \dots, T\}$  is associated with a multinomial  $\varphi_j$  over the  $W$ -word vocabulary, and each  $\varphi_j$  is drawn from a Dirichlet prior with parameter  $\beta$ . Similarly, each document  $d \in \{d_1, d_2, \dots, d_D\}$  is associated with a multinomial  $\theta_d$  over topics, drawn from a Dirichlet prior with parameter  $\alpha$ . The full procedure for generating a word in a document by LDA is as follows [29].

Each document in the corpus is a sequence of  $N_d$  words and the total corpus length is  $N$ . The procedure in Figure 3 implies a joint distribution over the random variables  $(w, z, \varphi, \theta)$ , which is given by equation (1)

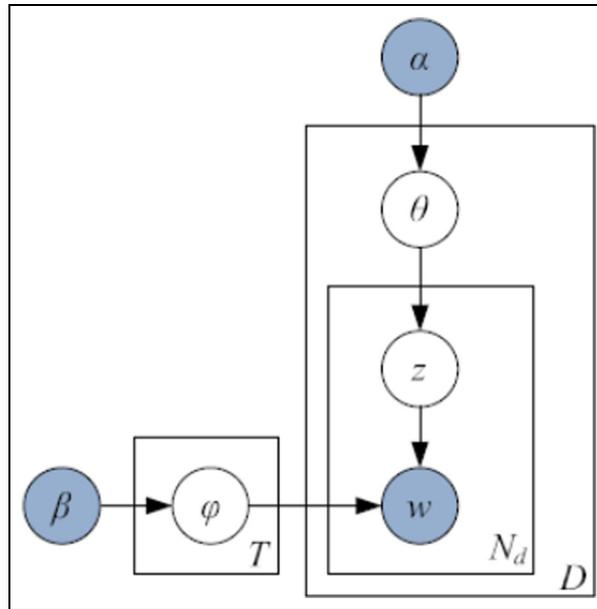


Figure 2. Graphical representation of the LDA model.

1. For each topic  $j \in \{1, \dots, T\}$ ,
  - a. Draw a topic-word multinomial distribution  $\varphi_j \sim \text{Dirichlet}(\beta)$ ;
2. For each document  $d \in \{d_1, d_2, \dots, d_D\}$ 
  - a. Draw a document-topic multinomial distribution  $\theta_d \sim \text{Dirichlet}(\alpha)$ ;
  - b. For each word  $w_i$  in document  $d$ ,
    - i. Draw a topic  $z_i \sim \text{Mult}(\theta_d)$ ;
    - ii. Draw a word  $w_i \sim \text{Mult}(\varphi_{z_i})$ ;

Figure 3. Definition of generative process in LDA.

$$\begin{aligned}
 P(w, z, \varphi, \theta | \alpha, \beta, d) &\propto \left( \prod_t^T p(\varphi_t | \beta) \right) \\
 &\times \left( \prod_j^D p(\theta_j | \alpha) \right) \left( \prod_i^N \varphi_{z_i}(w_i) \theta_{d_i}(z_i) \right)
 \end{aligned} \tag{1}$$

where  $\varphi_{z_i}(w_i)$  is the  $w_i$ th element in vector  $\varphi_{z_i}$ , and  $\theta_{d_i}(z_i)$  is the  $z_i$ th element in vector  $\theta_{d_i}$ . The goal of LDA is to find a set of model parameters, topic proportions and topic-word distributions. Standard statistical techniques can be used to invert the generative process of LDA, inferring the set of topics that were responsible for generating a collection of documents. The exact inference in LDA is generally intractable, therefore approximate inference algorithms are needed for posterior estimation [29,31,32]. The most common approaches that are used for approximate inference are expectation–maximisation, Gibbs sampling and variational method [29,32,42,43].

## 4. SAM: joint sentiment-aspect detection model

### 4.1. SAM description

LDA is a hierarchical model that treats data as arising from a generative process on random variables to extract the topics. SAM is an extension of LDA that uses a mixture of multinomials to model both topics and sentiments. In order to model sentiments, we propose SAM as a joint sentiment-aspect model by adding a sentiment layer to the LDA model, where sentiment labels are associated with documents, aspects are associated with sentiment labels and words are united

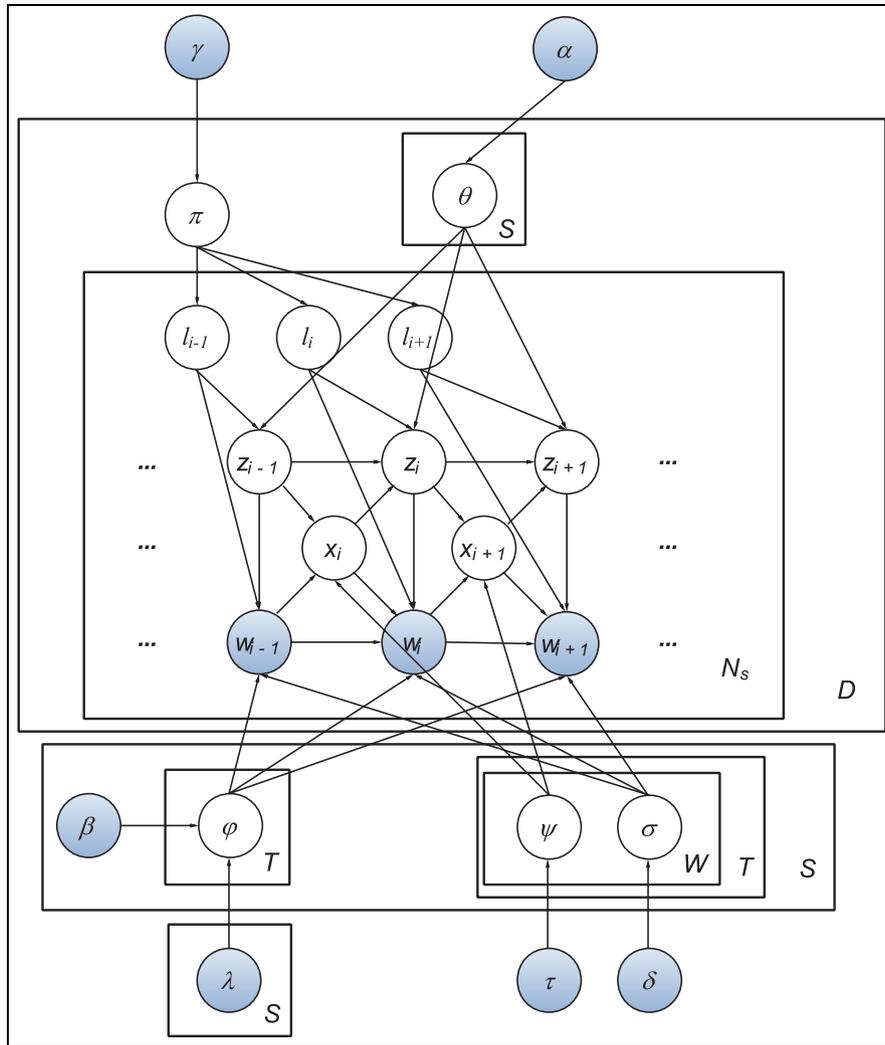


Figure 4. Graphical representation of SAM.

with both sentiment labels and aspects. SAM tries to extract latent aspects and sentiments simultaneously from reviews using word co-occurrences, frequencies and the order of words in each document. The proposed unsupervised approach can easily be transformed between domains or languages, where it can detect the polarity of text data at document, sentence or aspect level.

The proposed model tries to jointly extract latent sentiment-aspects from reviews by making use of information of documents as well as the order of words in each document. SAM model is similar to the LDA model in tying together parameters of different documents via a hierarchical generative model, but unlike the LDA model, it does not assume documents are ‘bag of words’ [29,40]. In other words, in LDA, the positions of individual words are neglected for topic inference, while SAM assumes that the topics of words in a document form a Markov chain, and that subsequent words are more likely to have the same topic or sentiment. Figure 4 shows the graphical representation of the SAM.

As shown in Figure 4, SAM is a four-layer hierarchical model, where sentiment labels are defined for review documents, aspects are associated with sentiment labels and the aspect of previous word, and words are related to sentiment labels and aspects. Assume that we have a dataset collection of  $D$  review documents denoted by  $\{d_1, d_2, \dots, d_D\}$ , each review in the dataset is a sequence of  $N_s$  sentences and  $N_d$  words, where each word in the sentence of the document is an entity from a vocabulary with  $W$  distinct words. Also  $S$  is the number of sentiment labels and  $T$  is the number of aspects. In this formal definition,  $S$  can be assigned to any number, where in the experiments, we used  $S = 2$  as for positive and negative sentiment labels. Therefore, the generative process of words in the SAM corresponding to the graphical representation shown in Figure 4 is as Figure 5.

1. For each sentiment label  $l \in \{1, \dots, S\}$ ,  
For each aspect  $j \in \{1, \dots, T\}$ ,
  - a. Draw a multinomial word distribution  $\varphi_{lj} \sim \text{Dirichlet}(\lambda_l \times \beta_{lj}^T)$ ;
2. For each sentiment label  $l \in \{1, \dots, S\}$ ,  
For each aspect  $j \in \{1, \dots, T\}$ ,  
For each word  $w_i \in \{1, 2, \dots, W\}$ ,
  - (a) Draw multinomial distribution  $\psi_{lji} \sim \text{Dirichlet}(\tau)$ ;
  - (b) Draw binomial distribution  $\sigma_{lji} \sim \text{Beta}(\delta)$ ;
3. For each document  $d \in \{d_1, d_2, \dots, d_D\}$ 
  - (b) Draw the document's sentiment distribution  $\pi_d \sim \text{Dirichlet}(\gamma)$ ;
  - (c) For each sentiment label  $l \in \{1, \dots, S\}$ ,
    - a. Draw an aspect distribution  $\theta_{dl} \sim \text{Dirichlet}(\alpha)$ ;
  - (d) For each sentence  $s \in \{1, \dots, N_s\}$  in document  $d$ ,  
For each word  $w_i$  in sentence  $s$ ,
    - i. Choose a sentiment label  $l_i^s \sim \text{Mult}(\pi_d)$
    - ii. Choose a status variable  $x_i^s \sim \text{Bin}(\psi_{l_i^s z_{i-1}^s w_{i-1}^s)$ ;
    - iii. If  $(x_i^s = 0)$  choose an aspect  $z_i^s \sim \text{Mult}(\theta_{dl_i^s})$ ;  
else  $z_i^s = z_{i-1}^s$ ;
    - iv. If  $(x_i^s = 1)$  choose a word  $w_i^s \sim \text{Mult}(\sigma_{l_i^s z_i^s w_{i-1}^s)$ ;  
else choose a word  $w_i^s \sim \text{Mult}(\varphi_{l_i^s z_i^s})$ ;

**Figure 5.** Formal definition of the generative process in SAM.

SAM, in addition to the two sets of random variables  $z$  and  $w$ , introduces two new sets of variables  $l$  and  $x$ , where  $l$  denotes the sentiment label of word  $w$  and  $x$  is defined to detect multi-word aspects from reviews. SAM assumes that the aspects in a sentence form a Markov chain with a transition probability that depends on  $\theta$ , sentiment label  $l_i$ , a random variable  $x_i$  and the aspect of previous word  $z_{i-1}$ . Random variable  $x$  denotes whether a multi-word aspect can be formed with the previous term or not. Therefore, SAM has the power to decide whether to assign the same aspect to subsequent words or to generate a unigram aspect, a bigram, a trigram and so on. If the model sets  $x_i$  equal to one, it means that  $w_{i-1}$  and  $w_i$  form a multi-word aspect, and if it is equal to zero, they do not. The notations of SAM are explained in Table 1.

With defining new sets of variables  $l$  and  $x$ , SAM assumes that each word  $w$  is associated with a latent topic (aspect)  $z$  and sentiment label  $l$ . Each of these aspects  $j \in \{1, \dots, T\}$  is associated with a multinomial  $\varphi_{lj}$  over the  $S$  sentiment labels and  $W$ -word vocabulary, and each  $\varphi_{lj}$  is drawn from a Dirichlet prior with parameters  $\beta$  and  $\lambda$ . Similarly, each document  $d \in \{d_1, d_2, \dots, d_D\}$  is associated with multinomial distributions  $\pi_d$  over sentiments and  $\theta_{dl}$  over sentiments and aspects, drawn from Dirichlet priors with parameters  $\gamma$  and  $\alpha$ , respectively. In addition to LDA, in SAM, two distributions  $\psi_{lji}$  and  $\sigma_{lji}$  are defined for each aspect  $j \in \{1, \dots, T\}$  and sentiment label  $l \in \{1, \dots, S\}$  over the  $i$ th word from  $W$ -word vocabulary.

The procedure in Figure 5 for generating a word  $w_i$  in document  $d$  under SAM results in three stages. First, one chooses a sentiment label  $l$  from the distribution  $\pi_d$ , then by the sentiment label  $l$  and aspect of previous word, one chooses status variable  $x_i^s$  from the binomial aspect-word distribution  $\psi_{lzw}$  over the sentiment label  $l$ . Following that, if the status variable  $x_i^s$  is zero, one chooses an aspect  $z_i^s$  from aspect distribution  $\theta_{d,l}$ , where  $\theta_{d,l}$  is conditioned on the

**Table 1.** Notations used in this article.

Concept	Symbol	Description
Data	$D$	Number of review documents
	$W$	Number of distinct words (vocabulary size)
	$N_d$	Number of words in document $d$
	$N_s$	Number of sentences in document $d$
	$d$	Document
SAM	$w$	Word
	$S$	Number of sentiment labels
	$T$	Number of topics
	$\alpha$	Dirichlet prior for $\theta$
	$\beta$	Dirichlet prior for $\varphi$
	$\delta$	Beta prior for $\sigma$
	$\tau$	Dirichlet prior for $\psi$
	$\lambda$	Parameter for transformation matrix
	$\gamma$	Dirichlet prior for $\pi$
	$\varphi$	Multinomial unigram sentiment-aspect distribution of words
	$\theta$	Multinomial distribution of aspects
	$\psi$	Binomial distribution of status variables
	$\sigma$	Multinomial bigram sentiment-aspect distribution of words
	$\pi$	Multinomial distribution over sentiments
	$z_i$	Latent aspect of $i$ th word in the document
	$l_i$	Latent sentiment label of $i$ th word in the document
	$x_i$	Bigram status between $(i-1)$ th word and $i$ th word
Model inference	$z_{-i}$	Aspect assignment for all words except word $i$
	$w_{-i}$	Vector of all words except word $i$
	$x_{-i}$	Status vector of all words except word $i$
	$l_{-i}$	Vector of sentiment label assignments for all words except word $i$
	$n_{lzw}$	Number of times word $w$ is assigned to aspect $z$ and sentiment label $l$ as a unigram
	$m_{lzw}$	Number of times word $v$ is assigned to topic $z$ and sentiment label $l$ as the second term of a bigram, given the previous word $w$
	$p_{zkw}$	Number of times the status variable $x = k$ , given the previous word $w$ and the previous word's aspect $z$
	$q_{dlz}$	Number of times a word assigned to aspect $z$ and sentiment label $l$ in document $d$
	$r_{dl}$	Number of times a word assigned to sentiment label $l$ in document $d$
	$\Gamma$	Gamma function

SAM: sentiment-aspect detection model.

sampled sentiment label  $l$ . Otherwise, if the status variable  $x_i^s$  is one,  $z_i^s$  is drawn from  $z_{i-1}^s$ . It is noteworthy that the aspect distribution of SAM is different from LDA, JST and ASUM. In LDA, there is only one aspect distribution for each review document, while SAM defines  $S$  different aspect distributions for each document. Also, SAM differs from JST and ASUM in that these two models use the aspect distribution  $\theta_d$  to sample an aspect, while SAM uses a conditioned rule and different distributions to draw an aspect. Finally, one draws a word from the word distribution conditioned on status variable  $x_i^s$ , aspect assignment  $z_i^s$  and sentiment label  $l_i^s$ . This is again different from LDA, JST and ASUM in that LDA samples a word only conditioned on aspects, and JST and ASUM only conditioned on aspect and sentiment label.

The hyperparameters  $\alpha$  and  $\beta$  in SAM can be treated as the prior observation counts for the number of times an aspect  $j$  with sentiment label  $l$  is sampled in a document and the number of times words sampled from single-word aspect  $j$  associated with sentiment label  $l$ , respectively, before having observed any actual words from that document [27,28]. Similarly, the hyperparameters  $\tau$  and  $\delta$  can be interpreted as the prior observation counts for the number of times a word form a multi-word aspect  $j$  with its previous word associated with sentiment label  $l$  and the number of times words sampled from a multi-word aspect  $j$  associated with sentiment label  $l$ , respectively, before any word from the corpus is observed. Also, Dirichlet parameter  $\gamma$  is the prior observation counts for the number of times sentiment label  $l$  sampled from a document before any observation from dataset collection. SAM is a weakly supervised model which uses a small seed set of sentiment seed words with no labelled training data. The parameter  $\lambda$  in SAM is a transformation matrix which modifies the Dirichlet priors  $\beta$  to capture the prior sentiment polarity of words. In other words, for each word  $w \in \{1, 2, \dots, W\}$  and the sentiment label  $l \in \{1, \dots, S\}$ , if  $w$  is found in the seed set, the value of  $\lambda_{l,w}$  is assigned as follows

$$\lambda_{l,w} = \begin{cases} 0.05 & \text{if } s_w = l \\ 0.95 & \text{otherwise} \end{cases}$$

where  $S_w$  is the prior sentiment label of word  $w$ , initially equal to zero for all words.

Based on the procedure in Figure 5, after model parameters have been determined, given a document  $d$ , the posterior probability of document  $d$  about the latent aspect, sentiment label and status variable of each word is defined as in equation (2)

$$P(z_i^d, x_i^d, l_i^d | w, z_{-i}^d, x_{-i}^d, l_{-i}^d, \alpha, \beta, \lambda, \gamma, \delta, \tau) = \frac{P(w_i^d, z_i^d, x_i^d, l_i^d | w_{-i}^d, z_{-i}^d, x_{-i}^d, l_{-i}^d, \alpha, \beta, \lambda, \gamma, \delta, \tau)}{P(w_i^d | w_{-i}^d, z_{-i}^d, x_{-i}^d, l_{-i}^d, \alpha, \beta, \lambda, \gamma, \delta, \tau)} \tag{2}$$

There are five sets of latent variables that we need to infer in SAM, that is, the per-document sentiment label-based aspect distribution  $\theta$ , the per-document sentiment distribution  $\pi$ , the per-collection single-word sentiment-aspect word distribution  $\varphi$ , the per-collection multi-word sentiment-aspect word distribution  $\sigma$  and the per-word sentiment label-based distribution of status variables  $\psi$  with regard to previous aspect and word.

Before discussing the inference problem of our model, it is worth noting that for increasing the consistency of the model, we could add to more sets of associations

$$\begin{cases} l_{i-1} \rightarrow l_i \\ x_i \rightarrow l_i \end{cases}$$

These two new associations define transition probabilities between sentiment labels and status variables, which means that a word and its previous could have the same sentiment label. Hence, a word has the option to inherit a sentiment label assignment from its previous word. Accordingly, we could substitute step 3(d)(i) in the generative process of Figure 5 as follows

$$\begin{aligned} &\text{draw } l_i^d \text{ from } \pi_d \text{ if } x_i^d = 0 \\ &\text{else } l_i^d = l_{i-1}^d \end{aligned}$$

### 4.2. Model inference

In order to estimate the distribution of  $\theta$ ,  $\pi$ ,  $\varphi$ ,  $\sigma$  and  $\psi$ , we use Gibbs sampling to estimate the posterior distribution over  $z$ ,  $l$  and  $x$  [48,49]. According to Gibbs sampling, each latent variable will be sequentially drawn with a probability distribution conditioned on current assignments for all other latent variables and the observed data [48]. Specifically, in SAM, the aspect assignment  $z_i^d$ , the sentiment label  $l_i^d$  and the bigram status  $x_i^d$  for each word token  $w_i^d$  can be drawn with the conditional probability distribution as in equation (3)

$$\begin{aligned} P(z_i^d, x_i^d, l_i^d | w, z_{-i}^d, x_{-i}^d, l_{-i}^d, \alpha, \beta, \lambda, \gamma, \delta, \tau) &= \frac{P(w_i^d, z_i^d, x_i^d, l_i^d | w_{-i}^d, z_{-i}^d, x_{-i}^d, l_{-i}^d, \alpha, \beta, \lambda, \gamma, \delta, \tau)}{P(w_i^d | w_{-i}^d, z_{-i}^d, x_{-i}^d, l_{-i}^d, \alpha, \beta, \lambda, \gamma, \delta, \tau)} \\ &\propto \left( \tau_{x_i^d} + p_{z_{i-1}^d, w_{i-1}^d, x_i^d} - 1 \right) \left( \gamma_{l_i^d} + r_{dl_i^d} - 1 \right) \\ &\times \begin{cases} \left( \alpha_{l_i^d, z_i^d} + q_{dl_i^d, z_i^d} - 1 \right) \frac{\beta_{w_i^d} + n_{l_i^d, z_i^d, w_i^d} - 1}{\sum_{v=1}^W \left( \beta_v + n_{l_i^d, z_i^d, v} \right) - 1} & \text{if } x_i^d = 0 \\ \left( \alpha_{l_i^d, z_{i-1}^d} + q_{dl_i^d, z_{i-1}^d} - 1 \right) \frac{\delta_{w_i^d} + m_{l_i^d, z_{i-1}^d, w_i^d} - 1}{\sum_{v=1}^W \left( \delta_v + m_{l_i^d, z_{i-1}^d, v} \right) - 1} & \text{if } x_i^d = 1 \end{cases} \end{aligned} \tag{3}$$

Letting the subscript  $-i$  denote a quantity that exclude data from  $i$ th index, the conditional probability distribution for  $z_i^d$ ,  $l_i^d$  and  $x_i^d$  can be equivalently separated as in equations (4)–(6)

$$P(z_i^d | w, z_{-i}^d, x, l, \alpha, \beta, \lambda, \gamma, \delta, \tau) \propto \begin{cases} \left( \alpha_{l_i^d, z_i^d} + q_{dl_i^d, z_i^d} - 1 \right) \frac{\beta_{w_i^d} + n_{l_i^d, z_i^d, w_i^d} - 1}{\sum_{v=1}^W \left( \beta_v + n_{l_i^d, z_i^d, v} \right) - 1} & \text{if } x_i^d = 0 \\ \left( \alpha_{l_i^d, z_{i-1}^d} + q_{dl_i^d, z_{i-1}^d} - 1 \right) \frac{\delta_{w_i^d} + m_{l_i^d, z_{i-1}^d, w_i^d} - 1}{\sum_{v=1}^W \left( \delta_v + m_{l_i^d, z_{i-1}^d, v} \right) - 1} & \text{if } x_i^d = 1 \end{cases} \tag{4}$$

$$P(x_i^d | w, z, x_{-i}^d, l, \alpha, \beta, \lambda, \gamma, \delta, \tau) \propto (\tau_{x_i^d} + p_{z_{i-1}^d w_{i-1}^d x_i^d} - 1) \times \begin{cases} \frac{\beta_{w_i^d} + n_{i_{z_i^d}^d w_i^d} - 1}{\sum_{v=1}^W (\beta_v + n_{i_{z_i^d}^d v}) - 1} & \text{if } x_i^d = 0 \\ \frac{\delta_{w_i^d} + m_{i_{z_i^d}^d w_{i-1}^d w_i^d} - 1}{\sum_{v=1}^W (\delta_v + m_{i_{z_i^d}^d w_{i-1}^d v}) - 1} & \text{if } x_i^d = 1 \end{cases} \quad (5)$$

$$P(l_i^d | w, z, x, l_{-i}^d, \alpha, \beta, \lambda, \gamma, \delta, \tau) \propto (\gamma_{l_i^d} + r_{dl_i^d} - 1) \times \begin{cases} \frac{\beta_{w_i^d} + n_{i_{z_i^d}^d w_i^d} - 1}{\sum_{v=1}^W (\beta_v + n_{i_{z_i^d}^d v}) - 1} & \text{if } x_i^d = 0 \\ \frac{\delta_{w_i^d} + m_{i_{z_i^d}^d w_{i-1}^d w_i^d} - 1}{\sum_{v=1}^W (\delta_v + m_{i_{z_i^d}^d w_{i-1}^d v}) - 1} & \text{if } x_i^d = 1 \end{cases} \quad (6)$$

The pseudo-code for Gibbs sampling procedure of SAM is shown in Figure 6.

In this method, we will initialise the count matrices by randomly assigning initial aspects, sentiment labels, the count matrices, and topic and sentiment assignment dictionaries. The Gibbs sampling procedure can be run until a stationary state of the Markov chain has been reached [48,49]. Markov chain samples are then used to approximate SAM parameters. The approximate probability of aspect  $z$  for sentiment  $l$  in review document  $d$ ,  $\theta_{lz}^d$ , is given in equation (7)

$$\theta_{lz}^d = \frac{\alpha_{lz} + q_{dlz}}{\sum_{t=1}^T (\alpha_{lt} + q_{dlt})} \quad (7)$$

The approximate probability of sentiment label  $l$  in review document  $d$  is shown in equation (8)

$$\pi_l^d = \frac{\gamma_d + r_{dl}}{\gamma + N_d} \quad (8)$$

The approximate probability of word  $w$  in a single-word aspect  $z$  and sentiment label  $l$  is given in equation (9)

$$\varphi_{lzw} = \frac{\beta_w + n_{lzw}}{\sum_{v=1}^W (\beta_v + n_{lv})} \quad (9)$$

The approximate probability of word  $v$  and sentiment label  $l$  in a multi-word aspect  $z$  given the previous word  $w$  is shown in equation (10)

$$\sigma_{lvw} = \frac{\delta_v + m_{lvw}}{\sum_{v=1}^W (\delta_v + m_{lvw})} \quad (10)$$

The approximate probability of the status variable  $x = k$  given the previous word  $w$  and the aspect  $z$  is given in equation (11)

$$\psi_{zwk} = \frac{\tau_k + p_{zwk}}{\sum_{k=0}^1 (\tau_k + p_{zwk})} \quad (11)$$

## 5. Experimental results

In this section, we describe the evaluation of the proposed SAM model in a variety of settings and compare it with the JST and ASUM models both qualitatively and quantitatively. For the qualitative analysis, we present a number of aspects and sentiments generated by all models including SAM to show that the SAM results are more informative, coherent and better correlated with the features of an object. For the quantitative comparison, the sentiment classification accuracy at document level for SAM was compared with the accuracy results for JST and ASUM for two different languages. All the experiments reported in this article have been carried out on a PC with an Intel(R) Core(TM) i7-2600 CPU at 3.40 GHz and 8GB memory running on a Windows 7 operating system. The application of the SAM model was programmed using Java programming language with J2SDK version 1.8.0 as the development environment. In the following sections,

```

Algorithm: Gibbs sampling procedure for SAM
Input: Corpus of review documents,  $\alpha, \beta, \delta, \tau, \gamma$ 
Output: Aspect and sentiment label assignment for all words in
the corpus
Method:
Initialize aspect and sentiment label assignment for each word in
the corpus of review documents with random sampling;
Initialize matrices  $\theta, \pi, \varphi, \sigma$  and  $\psi$  randomly;
While(stopping criterion of Gibbs sampling)
    For each review document  $d \in \{d_1, d_2, \dots, d_D\}$ 
        For each word  $w_i \in \{1, 2, \dots, W\}$ 
            Exclude word  $i$  associated with status value  $x$ , aspect
            assignment  $z$  and sentiment label  $l$  from counts  $p_{zwk}$ ,
             $q_{dlz}$ ,  $m_{lzwv}$ ,  $n_{lzw}$  and  $r_{dl}$ ;
            Sample a new status value, aspect assignment and
            sentiment label using equation 2;
            Update counts  $p_{zwk}$ ,  $q_{dlz}$ ,  $m_{lzwv}$ ,  $n_{lzw}$  and  $r_{dl}$  using the
            new status value, aspect assignment and sentiment
            label;
        End For
    End For
    For every 100 iterations
        Update matrices  $\theta, \pi, \varphi, \sigma$  and  $\psi$  with new sampling re-
        sults;
    End For
End While
End of Algorithm

```

**Figure 6.** Gibbs sampling procedure of SAM.

sentiment seed lexicon, sentiment classification, data collection, evaluation measure and the major experimental results will be discussed.

### 5.1. Sentiment classification

The main strategy of this article is to infer the semantic orientation of documents based on positive and negative polarity labels. This task is also known as document-level sentiment classification [20,35]. In the field of product review mining, with the results of document-level sentiment classification, users would know the necessary information to determine which products to purchase and companies would know the response from their customers and the performances of their competitors [20]. This article examines the accuracy of document-level sentiment classification of customer reviews as the quantitative analysis. Specifically, SAM uses majority vote of sentence-level sentiment counts, such that it utilises the sentiment distribution  $\pi$  of the sentences in a review to assign a sentiment label for the review.

### 5.2. Data collection

We deployed datasets of reviews for two different languages, English and Persian, and for three different domains, movie reviews, electronic device reviews and restaurant reviews. The English movie review dataset [35], *EMovies*, consists of

**Table 2.** Summary of review datasets.

Language	Dataset	Name	No. of reviews	Average no. of sentences/reviews
English	Movies	EMovies	2000	30
	Electronics	EElectronics	2000	10
	Restaurants	ERestaurants	1000	5
Persian	Electronics	PElectronics	1200	5

**Table 3.** Sentiment seed lexicon.

Negative	Awful, annoy, bad, complain, dangerous, disappoint, disturb, hate, horrible, inferior, nasty, negative, not good, not like, rubbish, sorry, unfortunate, waste, worthless, wrong
Positive	Attract, amazing, awesome, beautiful, best, comfort, correct, enjoy, excellent, fantastic, fortunate, good, great, like, love, nice, perfect, positive, satisfy, superior

**Table 4.** Contingency table.

Class	Predicted negative	Predicted positive
Negative	TN	FP
Positive	FN	TP

2000 reviews crawled from IMDB archive. For the English electronic review dataset, *EElectronics*, we used a collection containing 2000 reviews from Customer Reviews datasets [2,14], crawled from *Amazon.com* and *cnet.com*. Electronics dataset includes categories such as phone, digital camera, MP3 player and DVD player. For the English restaurant reviews [50], *ERestaurants*, we used 1000 reviews collected from Citysearch New York.<sup>1</sup> For the Persian reviews [21], we used a collection of 1200 electronic reviews, *PElectronics*, crawled from *www.digikala.com*. The summary of each dataset is given in Table 2. As can be seen from Table 2, other than the application and its related dictionary, the main difference between datasets in English language is the number of reviews and the average number of sentences per review.

### 5.3. Sentiment seed lexicon

We chose global opinion and evaluative words for the sentiment seed lexicon which can be treated as the paradigm for detecting the positive and negative sentiment orientation. These sentiment seed words are based on paradigms defined by Turney [20]. The seed lexicon consists of 20 positive and 20 negative sentiment words as shown in Table 3. The sentiment seed lexicon for Persian is the translated version of the English lexicon.

### 5.4. Evaluation measure

The performance of the SAM is evaluated using the accuracy measure, which has been used by several other researchers working on aspect-based sentiment analysis and sentiment classification [27,28]. Accuracy is the proportion of true results, both true positives (TPs) and true negatives (TNs), in the population. TP means that a sentence with positive polarity is classified by the system correctly. Accuracy is computed based on Table 4 as in equation (12)

$$Accuracy = \frac{TP + TN}{TN + FP + FN + TP} \quad (12)$$

where the sum ( $TN + FP + FN + TP$ ) represents all the sentences in a document.

### 5.5. Applying status variable

Based on the observation that aspects are nouns and sentiments are adjectives or adverbs [2], in the model, we examine the combination of noun phrases [23], adjectives and adverbs from review sentences. We use several experimentally

**Table 5.** Heuristic combinations of POS patterns.

Description	Patterns
Nouns	Bigram to four-gram of NN and NNS
Nouns and adjectives	Bigram to four-gram of JJ, NN and NNS
Determiners and adjectives	Bigram of DT and JJ
Nouns and verb gerunds	Bigram to trigram of DT, NN, NNS and VBG
Adjectives and adverbs	Bigram and trigram of JJ and RB

*I just bought the **digital camera** a few days ago. before I get used to it, here are my first feelings, the **picture quality** is so great, the **lens cover** is **surely loose**, the **zooming lever** is shaky, I hope it does not operate mechanically, otherwise you'll feel uneasy.*

**Figure 7.** Sample review.

extracted part-of-speech (POS) patterns, which we introduce as heuristic combinations in Table 5. In this article, we focus on five POS tags: NN, JJ, DT, NNS and VBG, for nouns, adjectives, determiners, plural nouns and verb gerunds, respectively.

From Table 5, the heuristic combinations of the first row select the candidate aspects from the noun phrase patterns such as ‘*NN NN*’ and ‘*NN NNS*’. The second row uses patterns such as ‘*JJ NN*’, ‘*JJ NNS*’ and ‘*JJ NN NN*’. The third row of Table 5 selects candidates based on the pattern ‘*DT JJ*’; the fourth row of the table uses heuristic patterns such as ‘*DT VBG*’, ‘*VBG NN*’ and ‘*NN VBG NN*’; and the last row selects patterns such as ‘*JJ JJ*’ for the sentiment word candidates. Based on the heuristic combinations of POS patterns presented in Table 5 for every word in each sentence, after the POS tagging assignments, the status variable  $x$  is assigned. The model sets  $x$  for the word equal to one when the two consecutive words form a multi-word aspect or a multi-word sentiment and sets to zero when they do not. Figure 7 is a portion of an actual review for a digital camera from www.amazon.com. The phrases in bold are examples for multi-word aspects and multi-word sentiments we intend to examine. For example, the phrase ‘*digital camera*’ is extracted from the pattern ‘*JJ NN*’, hence the status value of  $x$  for the word ‘camera’ is set to one. Also, the phrases ‘*picture quality*’ and ‘*lens cover*’ are based on the pattern ‘*NN NN*’ and the status values for the words ‘quality’ and ‘cover’ are set to one.

## 5.6. Comparative study

In our experiments, after preprocessing and extracting the sentences from the textual datasets, and after POS tagging, we used Gibbs sampling algorithm for all the models and ran the chain for 1000 iterations to produce a sample of latent variables for each of the experiments. Previous study has shown that topic models are not sensitive to hyperparameters and can produce reasonable results with a simple symmetric Dirichlet prior [27,39–43]. During Gibbs sampling, we used empirical values for the symmetric priors as given in Table 6. Distributions of each word in each aspect were then estimated by the models. We varied the number of extracted aspects in SAM, JST and ASUM for 1, 5, 10, 20, 50 and 80 aspects.

**5.6.1. Qualitative analysis.** For the qualitative analysis, examples of top extracted words of the models are presented in Tables 7 and 8. Table 7 shows the example of aspects detected by the models SAM, JST and ASUM under positive sentiment label from ERestaurant’s reviews, and Table 8 shows the results under negative sentiment label. In addition to the words, these tables present the probability of the word regarding the aspect assignment and sentiment label. In all models, the aspects seem to fit to the ERestaurants review data. For example, words such as ‘barbecued’, ‘codfish’, ‘desserts’, ‘garlic’, ‘spice’, ‘lunch’ and ‘service’ are related to the ERestaurant’s reviews and extracted by the models. In terms of aspect sentiment, by examining each of the aspects in Tables 7 and 8, it is quite evident that the aspects under the positive

**Table 6.** Hyperparameters setting.

Hyperparameters	Values
$\alpha$	1
$\beta$	0.01
$\delta$	0.01
$\tau$	0.1
$\gamma$	1

**Table 7.** Example of aspects detected by the models SAM, JST and ASUM under positive sentiment label from ERestaurant's reviews.

	Positive sentiment label					
	Aspect 1		Aspect 2		Aspect 3	
	Word	Probability	Word	Probability	Word	Probability
SAM	Barbecued	0.028	Food	0.048	Bagels	0.021
	Codfish	0.016	Great	0.010	Desserts	0.019
	Curiously	0.008	Dinner	0.007	Surprised	0.015
	Fabulous texture	0.007	Best	0.007	Pleasant	0.011
	Herb mix	0.007	Love	0.007	Service	0.008
JST	Garlic	0.030	View	0.057	Oil	0.029
	Opened	0.015	Boyfriend	0.049	Texture	0.029
	Tower	0.015	Spectacular	0.048	Stays	0.014
	Penne	0.015	Impressed	0.032	Thank	0.014
	Neighborhood	0.015	Magnificent	0.024	Spice	0.014
ASUM	Like	0.046	Cook	0.023	Service	0.031
	Fat	0.014	Dinner	0.016	Rarely	0.014
	Ladies	0.008	Style	0.011	Prices	0.013
	Lunch	0.008	Arguably	0.004	Imagine	0.012
	Oven	0.008	Common	0.003	Lined	0.010

SAM: sentiment-aspect detection model; JST: joint sentiment-topic model; ASUM: aspect-sentiment unification model.

and negative sentiment labels indeed bear positive and negative sentiments, respectively. For the words without sentiment, it can be found that these words appear in the context of both sentiment labels.

From the examples in Tables 7 and 8, it can be observed that SAM emphasises on extracting phrases and multi-word aspects along with unigram ones. Also for the assumption in SAM and ASUM, they extract more local topics and aspect-specific sentiment words. However, SAM overcomes ASUM shortcoming by considering word status and the structure of the sentences and positions of words in the documents.

**5.6.2. Quantitative analysis.** For the quantitative analysis, we compare the results of sentiment classification for the three models including SAM, JST and ASUM. All the models are evaluated using document-level sentiment classification accuracy. In SAM, a document is classified as positive if the positive sentiment sentence count is greater than negative sentences; otherwise, the document is labelled as negative. To determine the sentiment of a sentence, SAM uses sentiment distribution  $\pi$  of the sentence, such that the sentence is set to be positive if the probability of positive sentiment is higher or equal to the probability of negative sentiment, and vice versa. In contrast to SAM, JST and ASUM classify a document by viewing the document as a whole text and comparing the sentiment probabilities in the document level.

The models were run using different number of aspects, 1, 5, 10, 20, 50 and 80 in four datasets. When the aspect number is set to 1, all models, SAM, JST and ASUM, become very close to the standard LDA topic model with only  $S$  aspects, each of which represents a sentiment label. JST with one aspect is the same as LDA [27], but ASUM assumes that each sentence represents an aspect [27], whereas SAM assumes that consequent words may have the same aspect additionally. The accuracy results of sentiment classification are presented in Figures 8–11.

**Table 8.** Example of aspects detected by the models SAM, JST and ASUM under negative sentiment label from ERestaurant's reviews.

	Negative sentiment label					
	Aspect 1		Aspect 2		Aspect 3	
	Word	Probability	Word	Probability	Word	Probability
SAM	Slow	0.020	Dining	0.039	Restaurant	0.020
	Office	0.015	Food	0.022	Sitting	0.018
	Shockingly	0.011	Horrified	0.019	Worst	0.016
	Baby pizzas	0.010	Terrible management	0.016	Employees	0.011
	Dishes	0.010	Reputation	0.012	Expensive	0.009
JST	Conversation	0.048	Work	0.098	Took	0.096
	Routinely	0.024	Curry	0.049	Busy	0.048
	Servants	0.024	Look	0.032	Appetizer	0.048
	Paired	0.024	Stay	0.032	Tiny	0.024
	Hinting	0.024	Limited	0.016	Sangria	0.024
ASUM	Tourist	0.027	Pizza	0.032	Annoying	0.020
	Trap	0.021	Came	0.019	Scene	0.014
	People	0.010	Lot	0.014	Terrible	0.013
	Simply	0.009	Places	0.011	Location	0.012
	Worst	0.008	Pay	0.011	Oily	0.009

SAM: sentiment-aspect detection model; JST: joint sentiment-topic model; ASUM: aspect-sentiment unification model.

As can be seen from the figures, SAM has a significant improvement over other models in all the datasets and JST shows the lowest results among the models. Also, it can be observed in all three domains, where the number of aspects is equal to one, all the models are close together. This is because they act like classic LDA topic model.

From Figure 8, in the EElectronics, we can see that the sentiment classification accuracy for SAM, JST and ASUM increases when the number of aspects is getting large. Specifically, the JST model shows an accuracy value from 66.99% to 70.29% when the number of aspects changes from 1 to 10; the ASUM model attains an accuracy from 72.27% to 79.53% when the number of aspects changes from 1 to 20; and the SAM obtains the results from 72.27% to 86.79% when the number of aspects changes from 1 to 50. Similarly, for the ERestaurants, EMovies and PElectronics, we can see that the proposed model, SAM, has better performance over JST and ASUM. For the ERestaurants and PElectronics reviews, the curves are more smoothed, while in EMovies, all models attain the lowest results. Also, in EMovies, when the number of aspects increases, the model accuracy drops dramatically.

By assessing the results for EElectronics and PElectronics, it can be found that the results for EElectronics are slightly better. Using a complex script is the main challenge in Persian writing [21,51,52]. For example, one of the issues in Persian text mining is the wide variety of declensional suffixes. Another common problem of Persian text is word spacing. In Persian, in addition to white space as inter-words space, an intra-word space called pseudo-space separates word's part [21,51,52]. Using white space or not using any space instead of pseudo-space is a challenge in Persian reviews. For example, in the sentence 'این گوشی قابلیت تشخیص دستخط خوبی دارد' / 'This phone has good handwritten recognition ability', the word 'دستخط / handwritten' is a word which uses pseudo-space and contains two other words 'دست / hand' and 'خط / written line'. In this sentence, if the algorithm interprets the word 'دستخط / handwritten' united or separated, the feature space and the results could be different. These challenges affect the Persian sentiment classification accuracy [51].

By analysing Figures 8–11, the best sentiment classification accuracies of different settings for SAM, JST and ASUM models for four datasets are extracted in Table 9.

Table 9 shows each model in which the number of aspects obtains the best accuracy for the corresponding dataset. According to this table, comparing the domains, the highest performance among the models on average is found for the EElectronics review dataset. The reason is that topic model-based techniques perform better when the size of the training dataset is larger. The results of ERestaurants and PElectronics reviews are very close to EElectronics results, and the EMovies has the lowest results. When examining the datasets, it can be observed that the EMovies in size is the larger one but the poor results are because of many sentences in reviews that do not bear sentiment or opinion information. In Table 8, the lowest accuracy value is achieved by JST and the highest score is for SAM. Overall, from the results, it can be found that SAM outperforms other models in performance and that the aspects and sentiments generated by the SAM can significantly improve the performance over JST and ASUM.

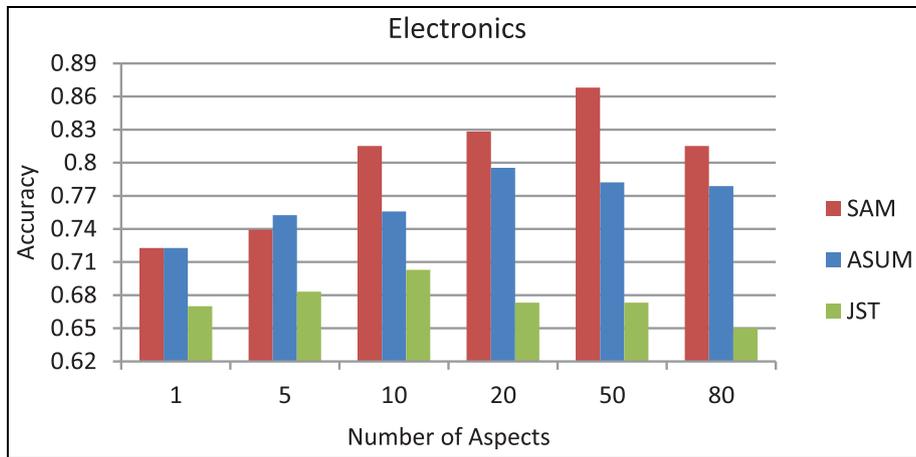


Figure 8. Sentiment classification accuracy results of SAM, JST and ASUM for EElectronics dataset.

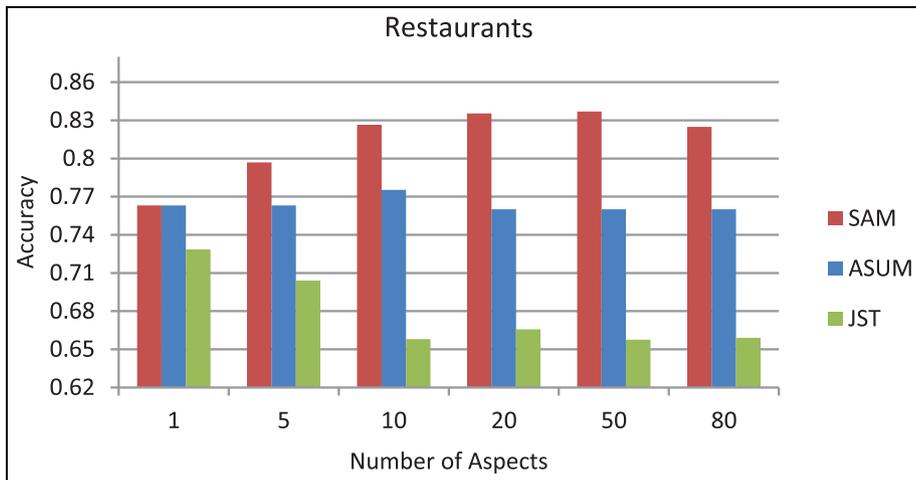


Figure 9. Sentiment classification accuracy results of SAM, JST and ASUM for ERestaurants dataset.

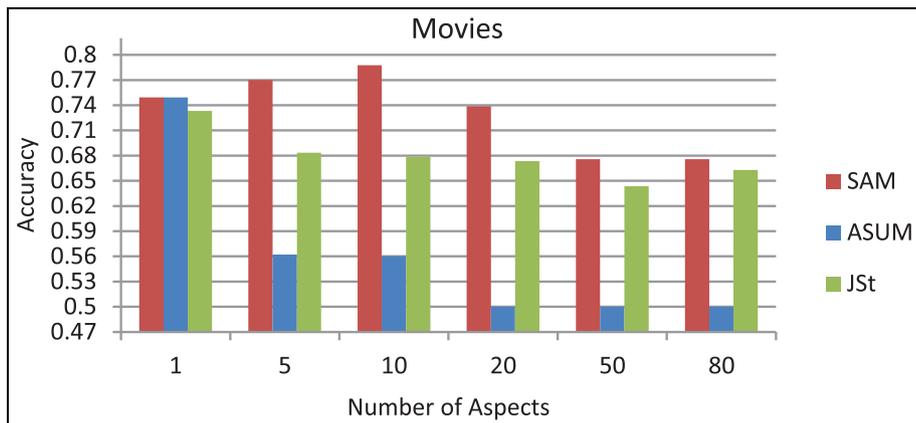
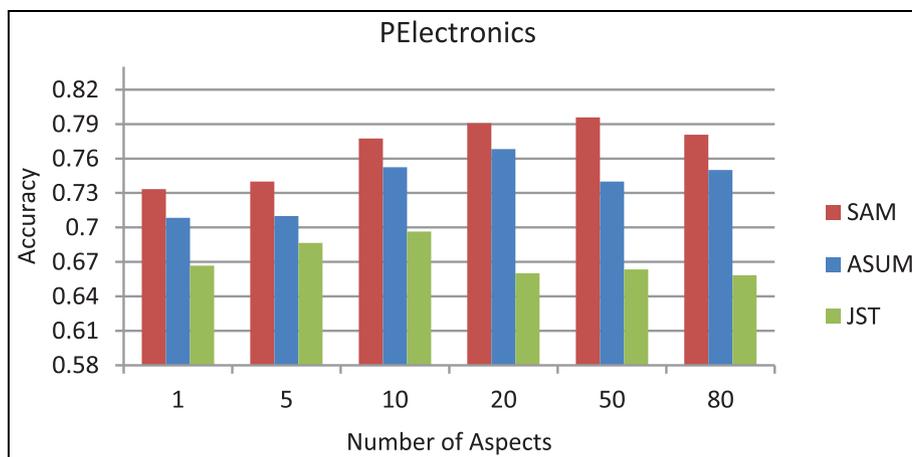


Figure 10. Sentiment classification accuracy results of SAM, JST and ASUM for EMovies dataset.



**Figure 11.** Sentiment classification accuracy results of SAM, JST and ASUM for PElectronics dataset.

**Table 9.** Best sentiment classification accuracy for each model in the datasets.

Dataset	SAM (no. of aspects)	JST (no. of aspects)	ASUM (no. of aspects)
EElectronics	86.79 (50)	70.29 (10)	79.53 (20)
ERestaurants	83.69 (50)	72.85 (1)	77.53 (10)
EMovies	78.77 (10)	73.31 (1)	74.94 (1)
PElectronics	79.58 (50)	69.63 (10)	76.83 (20)

SAM: sentiment-aspect detection model; JST: joint sentiment-topic model; ASUM: aspect-sentiment unification model.

### 5.7. Optimal number of aspects

From Table 9, it can be derived, on average, that when the number of aspects is between 10 and 50, the models obtain the best performance. Tables 10 and 11 show the sentiment classification accuracy for SAM, JST and ASUM in three domains with 10 and 50 aspects, respectively.

From Tables 10 and 11, it can be concluded that when comparing the models in 10 and 50 aspects, the optimum number of aspects for the SAM model is 50, where 10 aspects is the best number for JST and ASUM. The results with a different number of aspects show that 50 aspects per sentiment label capture more informative aspects and sentiments with few redundancies.

Finally, we can conclude that SAM for the problem of aspect-based sentiment analysis outperforms other models in all review datasets with multiple aspect settings. The bag-of-words assumption causes that the JST captures more global aspects from reviews, while ASUM assumes that each sentence represents an aspect which leads to a better model of aspect sentiment classification. In contrast, SAM releases the bag-of-words assumption by using word co-occurrences within sentence boundaries, assuming that each sentence represents one aspect and successive words may be from the same aspect to form a multi-word aspect. Therefore, SAM attains better effectiveness for detection of aspect and sentiment words.

This comparative evaluation shows that using an unsupervised approach for joint sentiment-aspect detection involving the assumption that the order of words matters, as well as the structure of review sentences and inter-relation information between words, and that giving more importance to multi-word aspects achieves promising performances.

## 6. Conclusion

In this article, we have presented an unsupervised language-independent model for an aspect-based approach towards the sentiment analysis task for review documents. This model, SAM, involves a probabilistic topic model which jointly detects aspects and sentiments from online reviews. SAM assumes that each sentence in a review contains an aspect and that consecutive words may form a phrase and have the same aspect. In other words, SAM is a probabilistic generative model which tries to detect aspects and sentiments from review documents by considering the underlying structure of a

**Table 10.** Sentiment classification accuracy in the datasets for each model with 10 aspects.

Dataset	SAM	JST	ASUM
EElectronics	81.51	70.29	75.57
ERestaurants	82.65	65.79	77.53
EMovies	78.77	67.85	56.03
PElectronics	77.75	69.63	75.24

SAM: sentiment-aspect detection model; JST: joint sentiment-topic model; ASUM: aspect-sentiment unification model.

**Table 11.** Sentiment classification accuracy in the datasets for each model with 50 aspects.

Dataset	SAM	JST	ASUM
EElectronics	86.79	67.32	78.21
ERestaurants	83.69	65.74	76.00
EMovies	67.56	64.35	50.00
PElectronics	79.58	66.33	74.00

SAM: sentiment-aspect detection model; JST: joint sentiment-topic model; ASUM: aspect-sentiment unification model.

document. The proposed approach models the aspect distributions with a Markov chain and releases the assumption that the aspect distribution within a document is conditionally independent. SAM differs from previous studies in that it needs no training labelled data: uses word status in review sentences, deals with detecting aspects and sentiments simultaneously and extracts coherent words for aspects and sentiments. Our experimental results indicate that SAM is effective in performing the task of sentiment classification and outperforms other models that combine aspect and sentiment detection such as JST and ASUM.

There are several ways that the research described here can be extended. One direction is to further improve and refine the proposed model by incorporating domain knowledge for aspect and sentiment detection. Another way of improvement is to apply statistical techniques for learning the parameters of the model. In other words, the model can be developed to extract aspect and sentiment jointly from review documents by choosing the hyperparameters automatically. Another variant is to evaluate sentiment analysis systems in online social media domains such as Twitter to investigate the performance for large-scale data processing. Finally, we aim to build a language-independent sentiment summarisation system.

### Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship and/or publication of this article.

### Funding

The author(s) received no financial support for the research, authorship and/or publication of this article.

### Note

1. <http://newyork.citysearch.com/>

### References

- [1] Liu B and Zhang L. A survey of opinion mining and sentiment analysis. In: Aggarwal CC and Zhai CX (eds) *Mining text data*. New York: Springer, 2012, pp. 415–463.
- [2] Hu M and Liu B. Mining and summarizing customer reviews. In: *Proceedings of the tenth ACM SIGKDD international conference on knowledge discovery and data mining*, Seattle, WA, 22–25 August 2004, pp. 168–177. New York: ACM.
- [3] Zhu J, Wang H, Zhu M et al. Aspect-based opinion polling from customer reviews. *IEEE T Affect Comput* 2011; 2(1): 37–49.
- [4] Liu B, Hu M and Cheng J. Opinion observer: analyzing and comparing opinions on the web. In: *Proceedings of the 14th international conference on World Wide Web*, Chiba, Japan, 10–14 May 2005, pp. 342–351. New York: ACM.

- [5] Brody S and Elhadad N. An unsupervised sentiment-aspect model for online reviews. In: *Proceedings of the Human Language Technologies: the 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, Los Angeles, CA, 2–4 June 2010, pp. 804–812. Stroudsburg, PA: Association for Computational Linguistics.
- [6] Wang C, Xiao Z, Liu Y et al. SentiView: sentiment analysis and visualization for internet popular topics. *IEEE T Hum Mach Syst* 2013; 43(3): 620–630.
- [7] Neviarouskaya A, Prendinger H and Ishizuka M. SentiFul: a lexicon for sentiment analysis. *IEEE T Affect Comput* 2011; 2(1): 22–36.
- [8] Bagheri A, Saraee M and de Jong F. An unsupervised aspect detection model for sentiment analysis of reviews. In: *Proceedings of the natural language processing and information systems*, Salford, 19–21 June 2013, pp. 140–151. New York: Springer.
- [9] Yu J, Zha ZA, Wang M et al. Aspect ranking: identifying important product aspects from online consumer reviews. In: *Proceedings of the 49th annual meeting of the Association for Computational Linguistics: Human Language Technologies*, Portland, OR, 19–24 June 2011, pp. 1496–1505. Cambridge, MA: MIT Press.
- [10] Qiu G, Liu B, Bu J et al. Opinion word expansion and target extraction through double propagation. *Comput Linguist* 2011; 37(1): 9–27.
- [11] Xu X, Cheng X, Tan S et al. Aspect-level opinion mining of online customer reviews. *Communications* 2013; 10(3): 25–41.
- [12] Thet TT, Na JC and Khoo CS. Aspect-based sentiment analysis of movie reviews on discussion boards. *J Inf Sci* 2010; 36(6): 823–848.
- [13] Popescu AM and Etzioni O. Extracting product features and opinions from reviews. In: *Proceedings of the conference on empirical methods in natural language processing*, Vancouver, BC, Canada, 6–8 October 2005, pp. 9–28. New York: ACM.
- [14] Bagheri A, Saraee M and de Jong F. Care more about customers: unsupervised domain-independent aspect detection for sentiment analysis of customer reviews. *Knowl-Based Syst* 2013; 52: 201–213.
- [15] Balahur A, Boldrini E, Montoyo A et al. Opinion question answering: towards a unified approach. In: *Proceedings of the 19th European conference on Artificial Intelligence*, Lisbon, 16–20 August 2010, pp. 511–516. Amsterdam: IOS Press.
- [16] Baccianella S, Esuli A and Sebastiani F. Multi-facet rating of product reviews. In: *Proceedings of the 31th European conference on IR research on advances in information retrieval*, Toulouse, 6–9 April 2009, pp. 461–472. Berlin: Springer.
- [17] Choi Y and Cardie C. Hierarchical sequential learning for extracting opinions and their attributes. In: *Proceedings of the 48th annual meeting of the association for computational linguistics*, Uppsala, 11–16 July 2010, pp. 269–274. Uppsala: Uppsala University.
- [18] Gerani S, Carman MJ and Crestani F. Proximity-based opinion retrieval. In: *Proceedings of the 33rd international ACM SIGIR conference on research and development in information retrieval*, Geneva, 19–23 July 2010, pp. 403–410. New York: ACM.
- [19] Titov I and McDonald R. A joint model of text and aspect ratings for sentiment summarization. In: *Proceedings of the Annual Conference Association for Computational Linguistics and the Human Language Technology (ACL-HLT)*, Columbus, OH: 15–20 June 2008, pp. 308–316. Stroudsburg, PA: Association for Computational Linguistics.
- [20] Turney PD. Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews. In: *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, Philadelphia, PA, 7–12 July 2002, pp. 417–424. New York: ACM.
- [21] Saraee M and Bagheri A. Feature selection methods in Persian sentiment analysis. In: *Proceedings of the conference natural language processing and information systems*, Salford, 19–21 June 2013, pp. 303–308. Berlin: Springer.
- [22] Soujanya P, Cambria E and Gelbukh A. Aspect extraction for opinion mining with a deep convolutional neural network. *Knowl Based Syst* 2016; 108: 42–49.
- [23] Bagheri A, Saraee M and de Jong F. ADM-LDA: an aspect detection model based on topic modelling using the structure of review sentences. *J Inf Sci* 2014; 40(5): 621–636.
- [24] Li F, Huang M and Zhu X. Sentiment analysis with global topics and local dependency. In: *Proceedings of the 24th Conference on Association of Advancement of Artificial Intelligence (AAAI)*, Atlanta, GA, 11–15 June 2010, pp. 1371–1376. Palo Alto, CA: AAAI Press.
- [25] Zhai Z, Liu B, Xu H et al. Constrained LDA for grouping product features in opinion mining. In: *Proceedings of the 15th Pacific-Asia conference on advances in knowledge discovery and data mining*, Shenzhen, 24–27 May 2011, pp. 448–459. New York: ACM.
- [26] Mei Q, Ling X, Wondra M et al. Topic sentiment mixture: modeling facets and opinions in weblogs. In: *Proceedings of the 16th international conference on World Wide Web*, Banff, AL, Canada, 8–12 May 2007, pp. 171–180. New York: ACM.
- [27] Lin C, He Y, Everson R et al. Weakly-supervised joint sentiment-topic detection from text. *IEEE T Knowl Data Eng* 2012; 24(6): 1134–1145.
- [28] Jo Y and Oh AH. Aspect and sentiment unification model for online review analysis. In: *Proceedings of the fourth ACM international conference on Web search and data mining*, Hong Kong, China, 9–12 February 2011, pp. 815–824. New York: ACM.
- [29] Blei DM, Ng AY and Jordan MI. Latent Dirichlet allocation. *J Mach Learn Res* 2003; 3: 993–1022.
- [30] Hofmann T. Probabilistic latent semantic indexing. In: *Proceedings of the 22nd annual international ACM SIGIR conference on research and development in information retrieval*, Berkeley, CA, 15–19 August 1999, pp. 50–57. New York: ACM.
- [31] Blei D. Introduction to probabilistic topic models. *Commun ACM* 2012; 55(4): 77–84.

- [32] Steyvers M and Griffiths TL. Probabilistic topic models. In: Landauer T, Mcnamara D, Dennis S et al. (eds) *Latent semantic analysis: a road to meaning*. Mahwah, NJ: Lawrence Erlbaum, 2005, pp. 424–440.
- [33] Titov I and McDonald R. Modeling online reviews with multi-grain topic models. In: *Proceedings of the 17th International Conference on World Wide Web*, Beijing, China, 21–25 April 2008, pp. 111–120. New York: ACM.
- [34] Yi J, Nasukawa T, Bunescu R et al. Sentiment analyzer: extracting sentiments about a given topic using natural language processing techniques. In: *Proceedings of the third IEEE international conference on data mining*, Melbourne, FL, 19–22 November 2003, pp. 427–434. New York: IEEE.
- [35] Pang B, Lee L and Vaithyanathan S. Thumbs up? Sentiment classification using machine learning techniques. In: *Proceedings of the empirical methods in natural language processing*, 6 July 2002, pp. 79–86. Stroudsburg, PA: Association for Computational Linguistics.
- [36] Moghaddam S and Ester M. ILDA: interdependent LDA model for learning latent aspects and their ratings from online product reviews. In: *Proceedings of the ACM SIGIR 34th international conference on research and development in information retrieval*, Beijing, China, 24–28 July 2011, pp. 665–674. New York: ACM.
- [37] Duc-Hong P and Le AC. Learning multiple layers of knowledge representation for aspect based sentiment analysis. *Data Knowl Eng* 2018; 114: 26–29.
- [38] Dasgupta S and Ng V. Topic-wise, sentiment-wise, or otherwise? Identifying the hidden dimension for unsupervised text classification. In: *Proceedings of the 2009 conference on empirical methods in natural language processing*, Singapore, 6–7 August 2009, pp. 580–589. New York: ACM.
- [39] Wallach HM. Topic modeling: beyond bag-of-words. In: *Proceedings of the 23rd international conference on machine learning*, Pittsburgh, PA, 25–29 June 2006, pp. 977–984. New York: ACM.
- [40] Gruber A, Weiss Y and Rosen-Zvi M. Hidden topic Markov models. In: *Proceedings of the eleventh international conference on artificial intelligence and statistics*, San Juan, 21–24 March 2007, pp. 163–170. PMLR (Proceedings of Machine Learning Research).
- [41] Wallach HM, Mimno D and McCallum A. Rethinking LDA: why priors matter. In: *Proceedings of the topic models: text and beyond workshop neural information processing systems*, 7–10 December 07–10 2009.
- [42] Wang X, McCallum A and Wei X. Topical N-grams: phrase and topic discovery, with an application to information retrieval. In: *Proceedings of the seventh IEEE international conference on data mining (ICDM 2007)*, Omaha, NE, 28–31 October 2007, pp. 697–702. New York: IEEE.
- [43] Andrzejewski D, Zhu X and Craven M. Incorporating domain knowledge into topic modeling via Dirichlet forest priors. In: *Proceedings of the 26th annual international conference on machine learning*, Montreal, QC, Canada, 14–18 June 2009, pp. 25–32.
- [44] Griffiths TL, Steyvers M and Tenenbaum JBT. Topics in semantic representation. *Psychol Rev* 2007; 114(2): 211–244.
- [45] Hai Z, Cong G, Chang K et al. Analyzing sentiments in one go: a supervised joint topic modeling approach. *IEEE T Knowl Data Eng* 2017; 29(6): 1172–1185.
- [46] Kaminka GA. A joint model for sentiment-aware topic detection on social media. In: *Proceedings of the ECAI 2016: 22nd European conference on artificial intelligence, The Netherlands – including prestigious applications of artificial intelligence (PAIS2016)*, The Hague, 29 August–2 September 2016, vol. 285, p. 338. Amsterdam: IOS Press.
- [47] Fu X, Yang K, Huang JZ et al. Dynamic non-parametric joint sentiment topic mixture model. *Knowl-Based Syst* 2015; 82: 102–114.
- [48] Resnik P and Hardisty E. *Gibbs sampling for the uninitiated*. Technical Report, UMIACS-TR-2010-04, University of Maryland, College Park, MD, 2010, <http://www.lib.umd.edu/drum/handle/1903/10058>
- [49] Minka T. *Estimating a Dirichlet distribution*. Technical Report, MIT, 2003, <https://tminka.github.io/papers/dirichlet/>
- [50] Ganu G, Elhadad N and Marian A. Beyond the stars: improving rating predictions using review text content. In: *Proceedings of the 12th workshop on web and databases*, 2009, <http://people.dbmi.columbia.edu/noemie/papers/webdb09.pdf>
- [51] Sadeghi M and Vegas J. Automatic identification of light stop words for Persian information retrieval systems. *J Inf Sci* 2014; 40(4): 476–487.
- [52] Sadeghi M and Vegas J. How well does Google work with Persian documents? *J Inf Sci* 2017; 43(3): 316–327.