Information Sciences 488 (2019) 190–204

Contents lists available at ScienceDirect

Information Sciences

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

Aspect based fine-grained sentiment analysis for online reviews

Feilong Tang[a,*], Luoyi Fu[a], Bin Yao[a], Wenchao Xu[b]

[a] *Department of Computer Science and Engineering, Shanghai Jiao Tong University, Shanghai 200240, China*
[b] *Department of Communication Engineering, East China Normal University, Shanghai 200241, China*

ARTICLE INFO

Article history:
Received 29 September 2018
Revised 24 February 2019
Accepted 25 February 2019
Available online 6 March 2019

Keywords:
Sentiment analysis
Online reviews
Topic model
Fine-grain aspects and opinions
Classifier

ABSTRACT

Fine-grained sentiment analysis for online reviews plays more and more important role in many applications. The key techniques here are how to efficiently extract multi-grained aspects, identify associated opinions, and classify sentiment polarity. Although various topic models have been proposed to process some of these tasks in recent years, there was little work available for effective sentiment analysis. In this paper, we propose a joint aspect based sentiment topic (JABST) model that jointly extracts multi-grained aspects and opinions through modeling aspects, opinions, sentiment polarities and granularities simultaneously. Moreover, by means of the supervised learning, we then propose a maximum entropy based JABST model (MaxEnt–JABST) to improve accuracy and performance in extracting opinions and aspects. Comprehensive evaluation results on real-world reviews for electronic devices and restaurants demonstrate that our JABST and MaxEnt–JABST models significantly outperform related proposals.

1. Introduction

With the rapid development of electronic commerce, it becomes more and more important to extract opinions and aspects of entities from various online reviews for both producers and consumers because these reviews are very opinion-rich [10]. Generally, consumers hope to firstly know which products are popular, and further which aspects of these products are high-quality, and finally why they are thought good or not. However, it is inefficient even impossible to analyze the above sentiment information manually due to massive review data and online processing requirement. So, aspect based sentiment analysis has been investigated in recent years, focusing on mining hidden knowledge automatically from online reviews.

In real world, different users are interested in different granularities of aspects. Aspect-specific opinion shows both sentiment polarity towards the aspect and explanation why the reviewer gave this opinion, which is commonly applied to extract aspects and opinions from online reviews. To sum up, to automatically extract aspects and corresponding opinions from online reviews, it is important to extract aspects, identify opinions associated the aspects, and classify sentiment polarity. For a fine-grained sentiment analysis, it is also important to identify granularities of aspects and opinions. Consequently, an efficient sentiment analysis has to explore aspects and opinions in multiple granularities. Recently, various researches based on topic models were proposed to process some of the above tasks. However, there is little work available to do all the tasks simultaneously so that they cannot use information from a task to improve results of other tasks, and also are incompetent for fine-grained sentiment analysis.

* Corresponding author.
E-mail addresses: tang-fl@cs.sjtu.edu.cn (F. Tang), fu-ly@cs.sjtu.edu.cn (L. Fu), yaobin@cs.sjtu.edu.cn (B. Yao), wchxu@ce.ecnu.edu.cn (W. Xu).

https://doi.org/10.1016/j.ins.2019.02.064
0020-0255/

Multiple granularity analysis is helpful to discover more information from online reviews. In this work, we consider how to analyze aspects and opinions in the *general* and *specific* granularities. General aspects and opinions refer to the basic ideas or background words included in most reviews, while specific ones are used to capture some fine-grained features so that aspect-specific opinions can catch more accurate information in reviews than single granularity sentiment analysis. The two-granularity analysis works in the following three steps. We describe this process in an example “This is a sleek, slim nice-looking computer. But using the touchpad is extremely frustrating”. Firstly, we extract aspects in this review, i.e., one general aspect “computer” and one specific aspect “touchpad”. Secondly, we identify three aspect-specific opinions: “nice-looking”, “sleek” and “slim” and one general opinion: “frustrating”. Finally, we classify and aggregate sentiments over extracted aspects. The “sleek”, “slim” and “nice-looking” on “computer” are positive opinions, while “frustrating” on “touchpad” is a negative opinion.

Some researchers have carried out investigations on the aspect-based sentiment analysis [21]. The representative schemes are multiple topic models to describe aspects and opinions. Along with this direction, Lin et al. extracted aspects and classified sentiment using a joint sentiment/topic (JST) model [9]. Jo in [5] proposed a unification model (ASUM) to find aspects and discover sentiment. Zhao et al. designed a maximum-entropy-based model ME-LDA [29] that separates global and specific opinions and aspects. HASM proposed by Kim et al. is a hierarchical aspect sentiment model [6]. Unfortunately, existing these work did not consider all the above tasks that are necessary to discover accurate sentiment.

In this paper, we propose a novel *joint aspect-based sentiment topic (JABST)* model that jointly captures multi-grain aspects and opinions. It tackles aspect extraction, opinion identification, sentiment polarity classification and separation of aspect and opinion words in a unified framework. In the JABST, the relationship among aspects, opinions, sentiment polarity and granularity is described using the latent variables in a graphical model. On the other hand, to improve the word distribution description, we propose another maximum entropy based model MaxEnt–JABST that extends JABST for better separating opinions and aspects.

The main contributions of this paper are summarized as follows.

- We identify the key tasks in fine-grained aspect-based sentiment analysis. A method for identifying granularities of aspect and opinion words is proposed to get a fine-grained sentiment analysis result.
- We propose a fine-grained aspect-based sentiment analysis model called joint aspect-based sentiment topic (JABST) model. This model extends topic models and is able to process key tasks of fine-grained aspect-based sentiment analysis simultaneously. JABST model can identify aspect and opinion words, extract aspect, analysis sentiment and is able to identify fine-grained aspects and opinions.
- We propose a sentiment analysis model MaxEnt–JABST based on maximum entropy. MaxEnt–JABST model is a semi-supervised version of fine-grained sentiment analysis model. It is able to better separate aspect and opinion words.
- We evaluate our JABST and MaxEnt–JABST models on reviews of electronic devices and restaurants qualitatively and quantitatively. The experimental results show that the proposed models outperform state-of-the-art baselines and are able to identify fine-grained aspects and opinions.

The remainder of this paper is organized as follows. Section 2 briefly reviews related work. We present our JABST and MaxEnt–JABST models in Section 3 and Section 4, respectively. In Section 5, we systematically evaluate our models. Finally, we conclude this paper in Section 6.

2. Related work

Joint aspect detection and sentiment analysis exhibit good performance in aspect-based sentiment analysis [21]. Syntax-based [30], machine learning [11,26] can be used for joint aspect and opinion identification. In this area, there have been many proposals.

2.1. Sentiment analysis based on aspects and opinions

Most researches on jointly extracting aspects and opinions used various topic models. Mei et al. proposed a TSM (topic sentiment mixture) model, handling background words, aspects and sentiments simultaneously [12]. But this model lacks explicitly among the sentiment and aspects. Improved from MG-LDA [26], MAS [27] added aspect ratings as an additional observed variable, which requires there is at least one rated aspect in one review. It is an impractical assumption. To extend LDA [1], Lin et al. proposed a JST (joint sentiment/topic) model [9] that adds a sentiment layer for more context-aware sentiment analysis. Lin et al. in [8] separated aspects and opinions based on sentiment lexiconto overcome the limitation of JST.

To simplify the sentiment analysis, Jo designed an aspect and sentiment unification model (ASUM) [5] that discovers pairs of aspect and sentiment in a sentence, under the assumption that one aspect and one sentiment can generate all words in each sentence. However, this assumption reduces its utilization, compared with the sliding window schemes [26] and the techniques injecting syntactic knowledge into topic models [7]. Moreover, its another limitation is that it is not able to handle fine-grained sentiment analysis. After that, maximum entropy [29] was used to enrich part-of-speech and lexicon information, based on the condition that one sentence has one aspect and conveys one opinion. Seeded aspect and sentiment

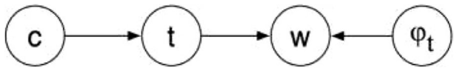

Fig. 1. An approach to identify fine-grained aspect and opinion words.

model [16] used seed words to find out aspect-based sentiment topics. Finally, Kim et al. proposed a hierarchical aspect sentiment model (HASM) [6] that describes the hierarchy of aspect-based sentiments using a tree. Similarly, this model does not consider the opinion and aspect granularity.

2.2. Topic models for aspect-based sentiment analysis

Aspect-based sentiment analysis schemes often jointly extract aspects and analyze feeing. Most of then used various topic models. Compared with the traditional support vector machine model, topic models have lower dimensions and exhibit lower computation complexity.

At the first stage, most topic models were designed based on latent semantic indexing (LSI) [20], and were extended into the PLSI (probabilistic latent semantic indexing) [3]. The latter is called as probabilistic latent semantic analysis (PLSA). Currently, there are some topic models concentrated on sentiment analysis with some other tasks. The leading topic model is latent dirichlet allocation (LDA) [1]. Although related, however, our models proposed a more generative and fine-grained approaches to aspect-based sentiment analysis. Focusing on review comments, [17] also used a maximum entropy classifier to separate topical and C-expression terms. Wang et al. [28] introduced a new aspect-based sentiment analysis task called latent aspect rating analysis, aiming to discover topical aspects with each individual reviewer's latent rating. In order to solve the same latent rating prediction problem, Moghaddam and Ester [14] proposed three different versions of topic models. Moghaddam and Ester [15] proposed the factorized LDA (FLDA) model to address the cold start issue, tackling items that have less than 10 reviews.

3. Aspect-based fine-grained sentiment analysis

In this section, we propose a fine-grained aspect-based sentiment analysis model called *joint aspect-based sentiment topic (JABST)* model, which models aspect, opinion, polarity and generality simultaneously. Our goal is to extract reasonable multi-grain aspect and opinion words from unlabeled online reviews. JABST models two distinct types of aspects and opinions: *general aspects, specific aspects, general opinions* and *aspect-specific opinions*.

3.1. Preliminaries

Sentiment analysis is a process or technique that identifies and extracts subjective information by means of analyzing and reasoning from given documents.

According to the handling granularity, sentiment analysis techniques fall into three categories: *document, sentence* and *aspect* levels. The document-level analysis is set for discovering general feeling to a document, with the results similar to product reviews. Sentence-level analysis focuses on the opinion of users to a given sentence, which can reveal user attitudes with more depth. Feature/aspect based sentiment analysis is set to mine users' opinions to different aspects.

Aspect based sentiment analysis is a fine-grained opinion mining process, which can explore the slight difference among interested entities. This approach can not only find the relationship among different aspect words but also can mine more accurate analysis results. To realize this goal a reality, we have to identify feature and opinion words, abstract these words for aspects, and finally mine these aspects for feelings.

JST (joint sentiment/topic model) is a representative scheme for aspect analysis, which sets up a relationship between opinions and entities described. Subsequently, an aspect and sentiment unification model (ASUM) was proposed to find aspects and opinions. Unfortunately, these two representative aspect analysis models all do not distinguish aspect and opinion words, as well as the granularity of aspects and opinions.

3.2. Key techniques in aspect-based sentiment analysis

The key point of aspect-based sentiment analysis is how to identify fine-grained aspects and opinions. Inspired by the existing sentiment analysis models, e.g., JST and AUSM, we add specified latent variables to represent the class and granularity of words, find the relationship among these variables and the words, and built the probability distribution of these words. Fig. 1 illustrates the framework that can efficiently recognize the fine-grained aspect and opinion words. Here, we use two variables c and t to represent classes of words and types of grained words, respectively. Specifically, c is an aspect word or an opinion word. The range of t is the Cartesian products of granularities and types. We use different word distribution variable φ_t to capture the different types of word distributions. This framework firstly determines the value of the class c and the type t, and then builds the word distribution φ_t, based on the classes of words.

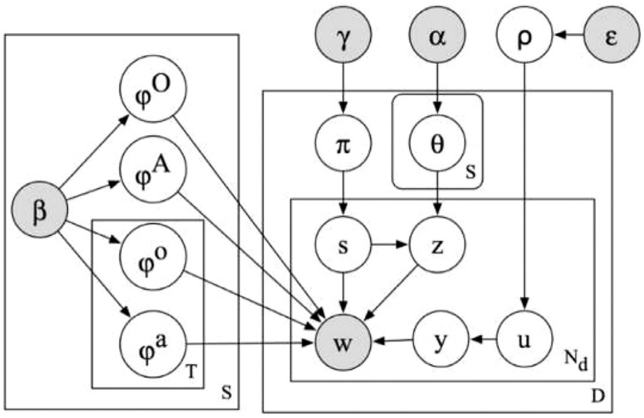

Fig. 2. Joint aspect-based sentiment topic (JABST) model.

Table 1
Notations for JABST model.

Plate notations	
S, D, T	the number of sentiment polarities, documents and topics, respectively
N_d	the number of words in a document *d*
R	the number of word subjectivity types
Hyper parameters	
α, β, γ	Dirichlet prior for θ, φ and π, respectively
ε	Beta prior for ρ
Latent & visible variables	
s, z, w	sentiment polarity, topic and word, respectively
y	the word type, including general and specific aspects, general and aspect-specific opinions
π	the multinomial distribution over sentiments
Θ	the multinomial distribution over topics
φ^O	the multinomial distribution over general opinion words
φ^A	the multinomial distribution over general aspect words
φ^a	the multinomial distribution over specific aspect words
φ^o	the multinomial distribution over aspect-specific opinion words
P	the binomial distribution over word subjectivity
U	the word subjectivity, which denotes whether it is an aspect or opinion word

3.3. JABST model

We propose a JABST model that can support the fine-grained sentiment analysis, which uses a latent variable to denote the word type and four word distributions, as shown in Fig. 2. The notations in the JABST model are listed in Table 1. In our four-layer JABST model, sentiment labels are associated with documents, topics are associated with sentiment labels, and words are associated with both sentiment labels and topics.

Fig. 2 illustrates a round of review generation in JABST model. Firstly, the JABST models determine all words in a review. Then, it finds out the subjective distribution of all the words, subsequent opinion distribution, and the topic distribution of each opinion. Next, the models explore the type of each word used and the corresponding distribution. Finally, he selects aspects and opinions. This process is repeated until the whole review is completely analyzed.

The focus of the JABST model is to find key probability distributions, including class distribution of words, opinion distribution of a document, subjective distribution of opinions, and aspect- and opinion-specific word distribution. These key probability distributions capture the aspects and opinions of reviewers, which will be presented in Section 3.3.2. In the following, we will present the detailed process of the sentiment analysis in the JABST model.

3.3.1. Sentiment generation in the JABST model

The JABST model generates sentiments in the following three steps. In step 1, JABST makes out a word subjectivity distribution $\rho \sim Beta(\varepsilon)$, a sentiment distribution $\pi_d \sim Dir(\gamma)$ for each document *d*, and a topic distribution $\theta_{d,s} \sim Dir(\alpha)$ for each sentiment *s* under a document *d*. In step 2, JABST firstly derives a general opinion distribution $\varphi_s^O \sim \mathrm{Dir}(\beta_s)$ and a general aspect distribution $\varphi_s^A \sim \mathrm{Dir}(\beta_s)$ for each sentiment *s*; and then derives a specific aspect distribution $\varphi_{s,z}^a \sim \mathrm{Dir}(\beta_s)$ and an aspect-specific opinion distribution $\varphi_{s,z}^o \sim \mathrm{Dir}(\beta_s)$ for a sentiment *s* under a topic *z*. In step 3, for each word *w* in a document *d*, *JABST* chooses a sentiment $s_i \sim Multi(\pi_d)$, a topic $z_i \sim Multi(\theta_{d,s})$, a subjectivity $u_i \sim B_i(\rho)$, a word type t_i based on a subjectivity u_i and a word $w_i \sim Multi(\varphi_{s_i,z_i}^{y_i})$ / $w_i \sim Multi\ (\varphi_{s_i}^{y_i})$.

Table 2
Notations for JABST model inference.

Notations	Description
V	vocabulary size
w_i, z_i, s_i	word in position i, topic of w_i, sentiment polarity of w_i
w, z, s	word list, assigned topics, sentiment polarities
z^{-i}, s^{-i}	topic assignment, sentiment polarity assignment except the one assigned to w_i
$n^{v,-i}$	the number of word v except w_i
n_d^{-i}	the number of words in document d except w_i
$n_{d,l}^{-i}$	the number of words assigned to sentiment l in document d except w_i
$n_{d,l,k}^{-i}$	the number of words assigned to sentiment l and topic k in document d except w_i
$n_{l,k}^{v,-i}$	the number of word v assigned to sentiment l and topic k except w_i
$n_{l,t}^{v,-i}$	the number of word v assigned to sentiment l and word type t expect w_i
$n_{l,k,t}^{v,-i}$	the number of word v assigned to sentiment l, topic k and word type t expect w_i
n_r^{-i}	the number of words assigned to word subjectivity r except w_i

The JABST model treats every word in one of four categories of roles: general aspect, specific aspect, general opinion and aspect-specific opinion, using the four word distributions: φ_s^O and φ_s^A refer to the general opinion and aspect word distribution under a sentiment s; $\varphi_{s,z}^o$ and $\varphi_{s,z}^a$ represent aspect-specific opinion and specific aspect word distribution for a sentiment s and a topic z.

We use a latent hidden variable $y \in \{0, 1, 2, 3\}$ to determine which distribution a word is derived from. In our model, we firstly separate opinion and aspect words, and then determine whether the word is general or specific. Moreover, we introduce a word subjectivity $u \in \{0, 1\}$, which is drawn from a binomial distribution ρ, to separate opinion and aspect words. In particular, a non-subjective word is an aspect word, and a subjective word is an opinion word describing a certain sentiment polarity.

3.3.2. Parameter estimation

There are multiple parameters in the JABST model. Parameter estimation is responsible for getting these parameters such as θ, π, φ and ρ. We use Gibbs sampling to estimate parameters for JABST model inference, which sequentially samples each interest variable, based on a sentiment lexicon for more accurate word classification. Specifically, in the beginning, we traverse all words in each document. If a word appears in the sentiment lexicon, it will be classified as an opinion word, i.e., $u=0$; otherwise, it will be classified as an aspect word such that $u=1$. Table 2 list the notations.

Parameter estimation uses a set of probability formulas. Our JABST samples interesting variables: the topic z_i, the sentiment polarity s, the subjectivity u_i and the type t_i of a word w_i. In the sampling process, the JABST model firstly extracts a topic k and a sentiment l and then extracts a subjectivity b and a word type t.

Specifically, during each round of the Gibbs sampling, the topic k and sentiment l of w_i are determined by the following conditional probability, formulated by (1)

$$P\left(z_i = k, s_i = l | \mathbf{z}^{-i}, \mathbf{s}^{-i}, \mathbf{w}, \alpha, \beta, \gamma\right) \propto \frac{n_{d,l}^{-i} + \gamma}{\sum_{l'}^{S} n_{d,l}^{-i} + S\gamma} \times \frac{n_{d,l,k}^{-i} + \alpha}{\sum_{k'}^{T} n_{d,l,k}^{-i} + T\alpha} \times \frac{n_{l,k}^{w_i,-i} + \beta}{\sum_{v}^{V} n_{l,k}^{v,-i} + V\beta} \tag{1}$$

The subjectivity b is selected based on the Beta distribution, using the probability formula in (2)

$$P(u_i = b) \propto \frac{n_b^{-i} + \varepsilon}{\sum_r^R n_r^{-i} + R\varepsilon} \tag{2}$$

Then, the word type t of w_i is sampled according to the conditional probability in the formula (3) such that

$$P(y_i = t) \propto \begin{cases} P\left(u_i = \frac{t}{2}\right) \cdot \frac{n_{l,t}^{w_i,-i} + \beta}{\sum_v^V n_{l,t}^{v,-i} + V\beta} & t = 0, 2 \\ P\left(u_i = \left\lfloor \frac{t}{2} \right\rfloor\right) \cdot \frac{n_{l,k,t}^{w_i,-i} + \beta}{\sum_v^V n_{l,k,t}^{v,-i} + V\beta} & t = 1, 3 \end{cases} \tag{3}$$

Based on the above probability, we can derive the approximate probability for the four important distributions: *the general opinion distribution* φ_s^O and *aspect word distribution* φ_s^A under a sentiment s, the *aspect-specific opinion distribution* $\varphi_{s,z}^o$ and *specific aspect word distribution* $\varphi_{s,z}^a$ for a sentiment s and a topic z as follows

$$\varphi_s^A = P(y = 0) = \frac{n_{u=0} + \varepsilon}{\sum_r^R n_r + R\varepsilon} \frac{n_{l,y=0}^{w_i} + \beta}{\sum_v^V n_{l,y=0}^{v} + V\beta} \tag{4}$$

$$\varphi_{s,z}^a = P(y = 1) = \frac{n_{u=0} + \varepsilon}{\sum_r^R n_r + R\varepsilon} \frac{n_{l,y=1}^{w_i} + \beta}{\sum_v^V n_{l,y=1}^{v} + V\beta} \tag{5}$$

Algorithm 1 Parameter estimation for the JABST model.

Input: document set $\{d_1, d_2, \ldots, d_D\}$, word sets of each document $\{w_1, w_2, \ldots, w_{Nd}\}$
Output: π, θ, φ and their corresponding matrixes Π, Θ 和 Ω
initialize Π of $D \times S$, Θ of $D \times S \times T$, Ω_A and Ω_O of $S \times V$, Ω_a and Ω_o of $S \times T \times V$
for $i = 1$ to D
  for $j = 1$ to N_d
    read $w_{i,j}$ from a document i
    compute aspect and opinion probability of $w_{i,j}$ using formula (1)
    generate $k_{i,j}$ and $l_{i,j}$ assigned to $w_{i,j}$
    compute subjective probability assigned to $w_{i,j}$ using formula (2)
    generate $b_{i,j}$ of $w_{i,j}$
    compute type probability assigned to $w_{i,j}$ using formula (3)
    generate the type $t_{i,j}$ of $w_{i,j}$
  end for
end for
update the matrix Π using formula (8)
update the matrix Θ using formula (9)
update matrixes Ω_A, Ω_a, Ω_O 和 Ω_o using formulas (4)–(7)

$$\varphi_s^O = P(y=2) = \frac{n_{u=1} + \varepsilon}{\sum_r^R n_r + R\varepsilon} \frac{n_{l,y=2}^{w_i} + \beta}{\sum_v^V n_{l,y=2}^{v} + V\beta} \tag{6}$$

$$\varphi_{s,z}^o = P(y=3) = \frac{n_{u=1} + \varepsilon}{\sum_r^R n_r + R\varepsilon} \frac{n_{l,y=3}^{w_i} + \beta}{\sum_v^V n_{l,y=3}^{v} + V\beta} \tag{7}$$

Finally, the probability of a sentiment l and a subject k can be formulated in (8) and (9), respectively

$$\pi = \frac{n_{d,l} + \gamma}{n_d + S\gamma} \tag{8}$$

$$\theta = \frac{n_{d,l,k} + \alpha}{n_{d,l} + T\alpha} \tag{9}$$

Algorithm 1 describes how to estimate parameters in the JABST model, including the joint sentiment/topic-document distribution θ, the joint sentiment/topic-word distribution φ^a and φ^o, the sentiment-word distribution φ^A and φ^O, and the sentiment-document distribution π. In particular, we first sample a topic k and a sentiment label l for each word at position i in each document d. Then we sample a word subjectivity b and word type t. Based on the multiple work distributions, we can derive top-n aspects and opinions that can capture users' interests with a high probability.

4. Fine-grained sentiment analysis based on the maximum entropy

JABST model essentially is an unsupervised learning mechanism. As pointed out in [9,12], unsupervised learning-based topic model cannot effectively distinguish opinions. Instead, supervised learning can significantly promote the extracting accuracy. In this Section, we proposed a more advanced model called *joint aspect-based sentiment topic model with maximum entropy (MaxEnt–JABST)* for more accurate sentiment analysis.

4.1. MaxEnt–JABST model

To improve the model ability in subjectivity description, we design the MaxEnt–JABST model that applies a maximum entropy classifier in the JABST to separate aspect and opinion words. MaxEnt–JABST model uses training results of the maximum entropy model in the generation process of topic models so that it combines the advantages of both the maximum entropy model and the topic model, as shown in Fig. 3. Here, λ_r refers to the maximum entropy model weights for x when value is r; x is a feature vector associated with w; other notations are the same as those in Table 1.

The key point in MaxEnt–JABST is how to derive the word subjectivity. Unsupervised method used in the JABST model can be used to separate aspects and opinions, which provides relatively coarse results. Since full unsupervised topic models cannot identify opinion words well, our MaxEnt–JABST model adopts supervised learning to estimate the subjectivity, i.e., separate aspects and opinions by means of syntactic information.

Two categories of features can be analyzed. One is the lexical features that include the previous, current and next words $\{w_{i-1}, w_i, w_{i+1}\}$; and another is the parts-of-speech (POS) tag features that include the previous current and next POS tags $\{\mathrm{POS}_{i-1}, \mathrm{POS}_i, \mathrm{POS}_{i+1}\}$ [29]. We prefer using POS tags to help discriminate between aspect and opinion words through a maximum entropy (ME) classifier. MaxEnt–JABST model divides the POS tags as 5 classifications, listed in Table 3.

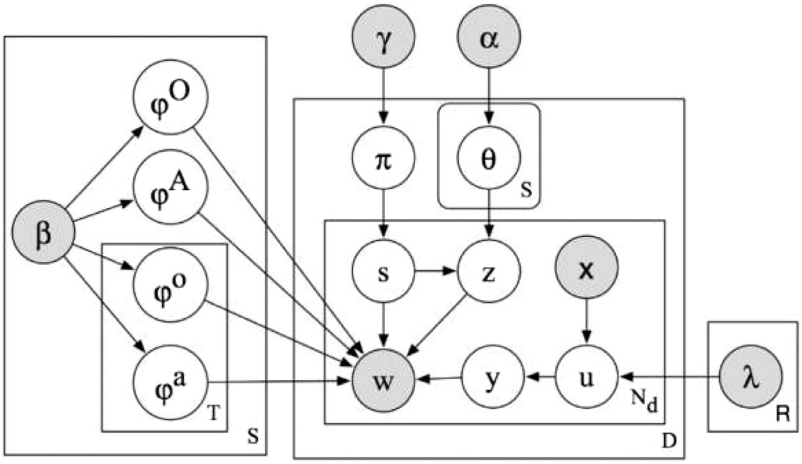

Fig. 3. MaxEnt–JABST model.

Table 3
Classifications for POS tags.

POS classifications	POS tags	Description
J	JJ, JJR, JJS	adj.
R	RB, RBR, RBS	adv.
N	NN, NNS, NNP, NNPS	noun.
V	VB, VBD, VBG, VBN, VBP, VBZ	verb.
o	Other tags	others

The sentiment generation process of the MaxEnt–JABST is similar to that in JABST, except for the step of drawing the word subjectivity distribution. In the MaxEnt–JABST, we train λ in advance using features associated with word subjectivity labels as the input of maximum entropy model. In the generative process, we draw the word subjectivity according the feature vector x and λ, where x means any arbitrary features that are effective and discriminative and λ refers to the maximum entropy weights. For a classification result c, we have the following weight

$$P(c = p|\mathbf{x}) = \frac{\exp(\lambda_p \cdot \mathbf{x})}{\sum_{p'} \exp(\lambda_{p'} \cdot \mathbf{x})} \tag{10}$$

4.2. Parameter estimation in MaxEnt–JABST model

MaxEnt–JABST model estimates parameters using the strategies similar to the above JABST model, besides adding the word subjectivity derivation. During each iteration, the sentiment and topic of w_i can be calculated in the formula (1). The maximum entropy classifier in MaxEnt–JABST determines the word subjectivity of w_i with the following probability

$$P(u_i = b) = \frac{\exp(l_b \cdot x_{w_i})}{\sum_r^R \exp(l_r \cdot x_{w_i})} \tag{11}$$

Based on the above formulas, we can have the four probability formulas of a word w as follows

$$\varphi_s^A = P(y=0) = \frac{\exp(l_{u=0} \cdot x_{w_i})}{\sum_r^R \exp(l_r \cdot x_{w_i})} \frac{n_{l,y=0}^{w_i} + \beta}{\sum_v^V n_{l,y=0}^v + V\beta} \tag{12}$$

$$\varphi_{s,z}^a = P(y=1) = \frac{\exp(l_{u=0} \cdot x_{w_i})}{\sum_r^R \exp(l_r \cdot x_{w_i})} \frac{n_{l,y=1}^{w_i} + \beta}{\sum_v^V n_{l,y=1}^v + V\beta} \tag{13}$$

$$\varphi_s^O = P(y=2) = \frac{\exp(l_{u=1} \cdot x_{w_i})}{\sum_r^R \exp(l_r \cdot x_{w_i})} \frac{n_{l,y=2}^{w_i} + \beta}{\sum_v^V n_{l,y=2}^v + V\beta} \tag{14}$$

$$\varphi_{s,z}^o = P(y=3) = \frac{\exp(l_{u=1} \cdot x_{w_i})}{\sum_r^R \exp(l_r \cdot x_{w_i})} \frac{n_{l,y=3}^{w_i} + \beta}{\sum_v^V n_{l,y=3}^v + V\beta} \tag{15}$$

The distributions that we need for further analysis are the same as (4). The Gibbs sampling in MaxEnt–JABST is similar to Algorithm 1, except using different probability formulas.

Table 4
Statistics of the data sets.

Data sets	Amazon	Yelp
Number of reviews	24,259	25,456
Number of reviews with 4+ stars	72%	68%
Average number of words per review	182	160
Average number of tokens preprocessed per review	87	76

5. Performance evaluations

We evaluate our JABST and MaxEnt–JABST models in terms of various performance metrics using real-world datasets, and analyze the impact of important parameters.

5.1. System settings

5.1.1. Parameters

As most schemes, we set common hyper parameters as $\alpha = 50$ / T, $\beta = 0.01$, $\gamma = 0.1$ and $\varepsilon = 0.5$; and set sentiment number $S=2$ to distinguish positive and negative opinions. In the experiments, we extract top-5 specific aspects and aspect-specific opinions and top-10 global aspect and opinion words, respectively. In the maximum entropy classifier training, we randomly selected 100 sentences in reviews from each dataset. These parameters were adopted in all the compared models. Finally, all models were trained 1000 iterations.

Moreover, the sentiment lexicon[1] in [4] was used as the prior information to assign word tokens to sentiment labels and topics initially. Specifically, the word token is assigned with the corresponding sentiment label if there is a match. Otherwise, a sentiment label is randomly sampled for a word token.

5.1.2. Compared models

The *JABST* model generates the word subjectivity using an unsupervised way while the *MaxEnt–JABST* using a Maximum Entropy classifier built on Maximum Entropy Toolkit.[2] We evaluated both *JABST and MaxEnt–JABST* models with the following representative work.

- *JST* [9], which is a joint sentiment/topic model. We downloaded the source code from the authors' Github.[3]
- *ASUM* [5], which is an aspect and sentiment unifications model. We downloaded the system from the authors' homepage.

5.2. Datasets

Our experiments were conducted on the two sets of online reviews[4] from [5]. The first one is a set of reviews on seven categories of electronic devices: air conditioner, canister vacuum, coffee machine, digital SLR, laptop, MP3 player and space heater from Amazon. Each category of data has around 5000 reviews. The second is a set of reviews on restaurants from Yelp, with about 25,000 reviews. These reviews were randomly selected from the reviews on the 320 most rated restaurants in Atlanta, Chicago, Los Angeles, and New York City.

The two datasets were pre-processed, removing URL, non-English alphabets and stop words, and then were tagged according to the Stanford POS Tagger.[5] To extract the contextual features, we also keep the data containing stop words to train our maximum entropy classifier. The statistics of pre-processed datasets are showed in Table 4.

5.3. Results and analysis

The focus of this work is multi-grain aspects and opinions extraction ability of our models. So, we firstly compared the quality of extracted topics, then evaluate the precision of aspects and opinions.

5.3.1. Qualitative evaluation

Our JABST and MaxEnt–JABST models extract words from word distributions. We show an example of extracted general aspect, opinion words and three examples of specific aspect, opinion words under positive and negative sentiment labels in Table 5 using Yelp dataset and Table 6 using Amazon dataset, respectively.

From these two tables, we find the general opinions can well describe the sentiment polarity; and the extracted topics are quite informative and coherent. For example, the second topic under positive sentiment praises the delicious food like

[1] https://www.cs.uic.edu/~liub/FBS/sentiment-analysis.html#lexicon.
[2] http://homepages.inf.ed.ac.uk/lzhang10/maxent_toolkit.html.
[3] https://github.com/linron84/JST.
[4] http://uilab.kaist.ac.kr/research/WSDM11.
[5] http://nlp.stanford.edu/software/tagger.shtml.

Table 5
Sampling aspects and opinions in the Yelp dataset using MaxEnt–JABST.

Sentiment	General				Topic1		Topic2		Topic3	
	Aspect		Opinion		Aspect	Opinion	Aspect	Opinion	Aspect	Opinion
Positive	food	flavor	best	nice	dessert	sweet	fri	satisfi	ramen	delici
	restaur	meal	great	special	chocol	amaz	burger	juici	soup	impress
	place	thing	right	well	flavor	rich	bun	homemad	pork	fresh
	one	menu	good	enjoy	cream	lol	chees	recommend	noodl	awesom
	order	look	like	enough	ice	cute	cook	actual	bowl	cheap
Negative	place	service	bad	sorri	tea	weird	burger	simple	meat	small
	food	us	hate	suck	water	small	fri	greasi	bowl	spici
	restaur	citi	problem	hard	ice	bore	beer	salti	rice	hot
	waiter	dessert	complaint	disappoint	coffe	wast	onion	gravi	ramen	challeng
	menu	would	oh	far	cup	crazi	waffl	wtf	beef	fat

Table 6
Sampling aspects and opinions in the Amazon dataset using MaxEnt–JABST.

Sentiment	General				Topic1		Topic2		Topic3	
	Aspect		Opinion		Aspect	Opinion	Aspect	Opinion	Aspect	Opinion
Positive	compani	area	good	favorit	pc	work	touch	avail	batteri	good
	purchas	amazon	great	excel	laptop	cool	show	want	life	long
	product	one	worth	fine	mac	reliabl	play	easili	charg	enough
	order	ship	nice	love	softwar	wireless	hd	free	last	nice
	price	laptop	like	well	support	solid	screen	clear	time	fit
Negative	product	machin	problem	disappoint	comput	crash	button	never	life	nois
	buy	money	bad	worri	laptop	problem	view	lock	batteri	littl
	use	brand	complaint	useless	window	frustrat	screen	difficult	netbook	slow
	qualiti	cofee	look	wrong	system	hard	menu	back	power	heavi
	one	hour	hard	regret	instal	slow	display	annoy	perform	extra

juicy burger and homemade cheese. And general aspect words are commonly concerned entities in the dataset domain, like food or restaurant. In our models, both aspect words that described the domain entity and background words will be assigned to general aspect, because they occurred across the reviews.

5.3.2. Sentiment classification accuracy

Accurate classification is the core of sentiment analysis from reviews. We evaluate the four models in terms of accuracy of sentiments mined from the reviews on Amazon products and restaurants in Yelp. Both the datasets use a 5-star rating system. In this part, the reviews, which have positive sentiments with higher probability than negative sentiment, are set as positive; otherwise, it is set to negative. We compare all the reviews that are not 3-stars to calculate sentiment classification precision. Specifically, aspects and opinions with 1–2 star rating and 4–5 star rating were identified as negative and positive ones, respectively. We set the number of topics as 10, 15, 20, 25 and 30 to evaluate the impact of topic numbers.

As shown in Fig. 4, both JABST and MaxEnt–JABST outperform the ASUM and JST. Generally, the accuracy increases as topics increase because the models better fit the data. JST also outperforms ASUM in Yelp dataset since our pre-processed data benefits to JST. MaxEnt–JABST improves average accuracy around 5% over JABST. As the number of topics increases, the improvement grows up faster and faster.

5.3.3. Topic extraction quality

The quality of extracted topics is important to measure topic extraction quality because it has a significant impact on aspect and opinion identification. We use the topic coherence to evaluate the topic extraction quality. Higher topic coherence is, higher the quality of the topic will be. Coherent topics extracted from reviews can provide specific semantics. It was well-known that both the intrinsic measurement UMass [13] and the extrinsic measurement UCI [18] align with human evaluations [22]. In this work, we only use the UMass to evaluate the four models since UCI usually requires an external corpus. In particular, we set the topic number as 10, 15, 20, 25 and 30, using top 10 words to calculate UMass score of topics. Then we compare the average of topic coherence from each dataset.

Fig. 5 illustrates that our JABST and MaxEnt–JABST models outperform the LDA, JST and ASUM, which results from the fact that our models can extract more accurate and interpretable topics in aspects and opinions. Particularly, MaxEnt–JABST exhibits a higher topic quality. So, maximum entropy priors are more efficient than Beta priors in opinion and aspect analysis.

5.3.4. Opinion and aspect identification

The number of topics was set as 20 for each sentiment. We use the method in [5] to find top aspect words and opinion words under a topic. For each topic under a certain sentiment, we manually judge if the presented top n aspect or opinion

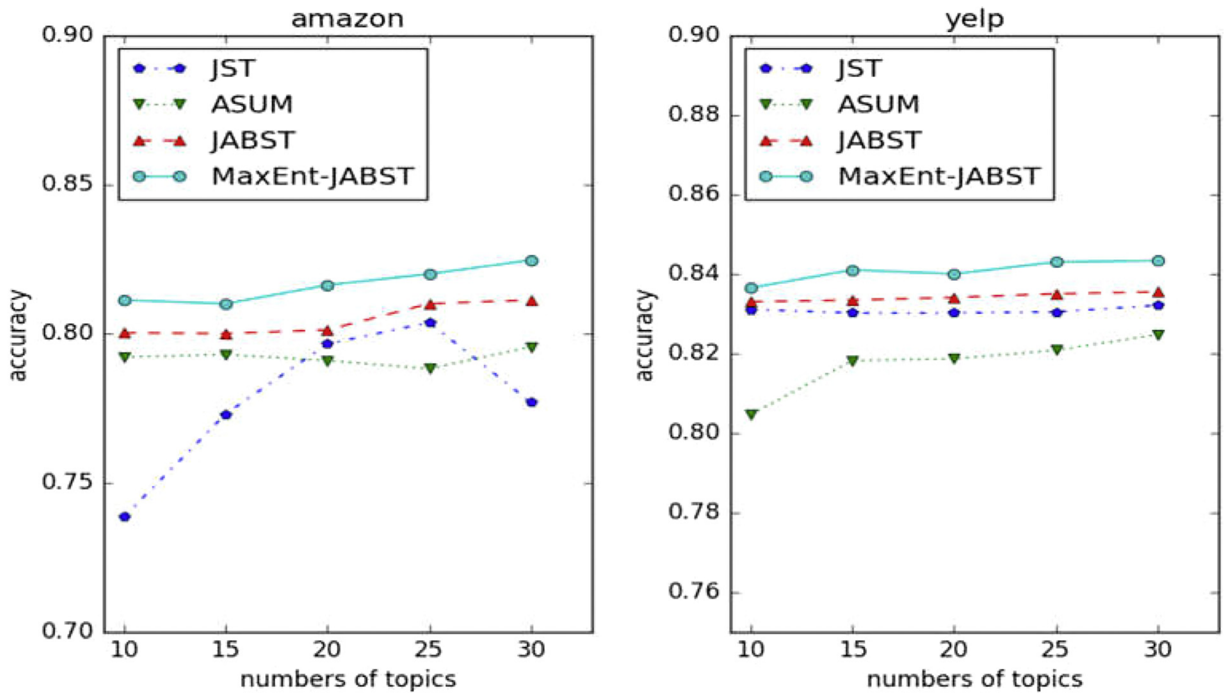

Fig. 4. Average polarity accuracy.

Table 7
Examples of specific opinions identified from ASUM, JABST and MaxEnt–JABST.

Topic 1			Topic 2		
MaxEnt–JABST	JABST	ASUM	MaxEnt–JABST	JABST	ASUM
larg	interest	*great*	loud	*hot*	*problem*
typic	light	friendli	noisi	exhaust	*bad*
uniqu	larg	*good*	*hot*	*cool*	*issu*
plenti	*local*	*nice*	slow	noisi	*like*
neat	new	*excel*	plastic	*lower*	*disappoint*
casual	*enjoy*	*realli*	awkward	*neg*	light
relax	*feel*	*alway*	dust	larg	noisi
fantast	uniqu	delici	exhaust	notsatisfi	low
happi	*variou*	*love*	*humid*	expens	*complain*
outstand	uncommon	*amaz*	leak	fault	*never*

words are correct. An aspect-specific opinion word is correct when it is (1) the opinion word under the sentiment polarity (2) reasonable describing the aspect. Then, we calculated the precision for this topic. We use precision@n, a common metric, for performance evaluation. We set n to 5 and 10 because top words are more important. Moreover, we random selected five topics for each sentiment, and calculated the average precision of all topics. The results in Fig. 6 shows that extracted opinions and aspects in our JABST and MaxEnt–JABST are more specific.

We list examples of the three models in two topics, where the opinions extracted mistakenly are marked with italic and red. From Table 7, we find that the ASUM model extracts many general opinions like bad, good and nice. Instead, our JASBT and MaxEnt–JABST can extract more specific opinion words that have more strong correlation with aspects, e.g., uniqu, dust and casual, and can explain sentiment analysis results more clearly. The results in Fig. 6 show that the opinion words in JABST and MaxEnt–JABST are more specific. According to the opinion precision result, the MaxEnt–JABST and JABST have average improvements of 18.9% and 16.7% over ASUM. On the other hand, we evaluated our models in aspect identification. We consider the aspect word is specific when it is the aspect word that describes the topic.

According to Table 8, we also can find that the ASUM generates more general words, e.g., one and place, while our JABST and MaxEnt–JABST generate more specific aspects like power, attach and so on. As a result, the MaxEnt–JABST and JABST improve 13.1% and 9.8% over ASUM in terms of precision, as shown in Fig. 7.

5.3.5. Feature selection

In this part, we evaluate how lexical feature and part of speech feature impact the feature selection. In the experiments, we used lexical features (i.e., w_{i-1}, w_i, w_{i+1}) and part of speech features (i.e., POS_{i-1}, POS_i, POS_{i+1}) feature vectors to train

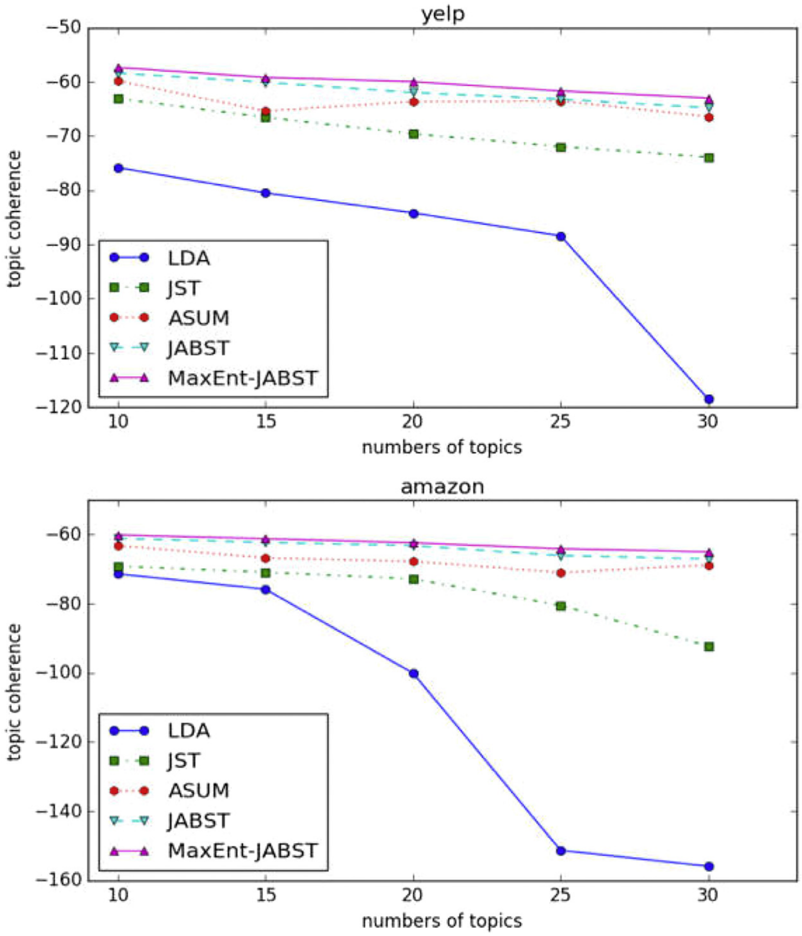

Fig. 5. Average topic coherence.

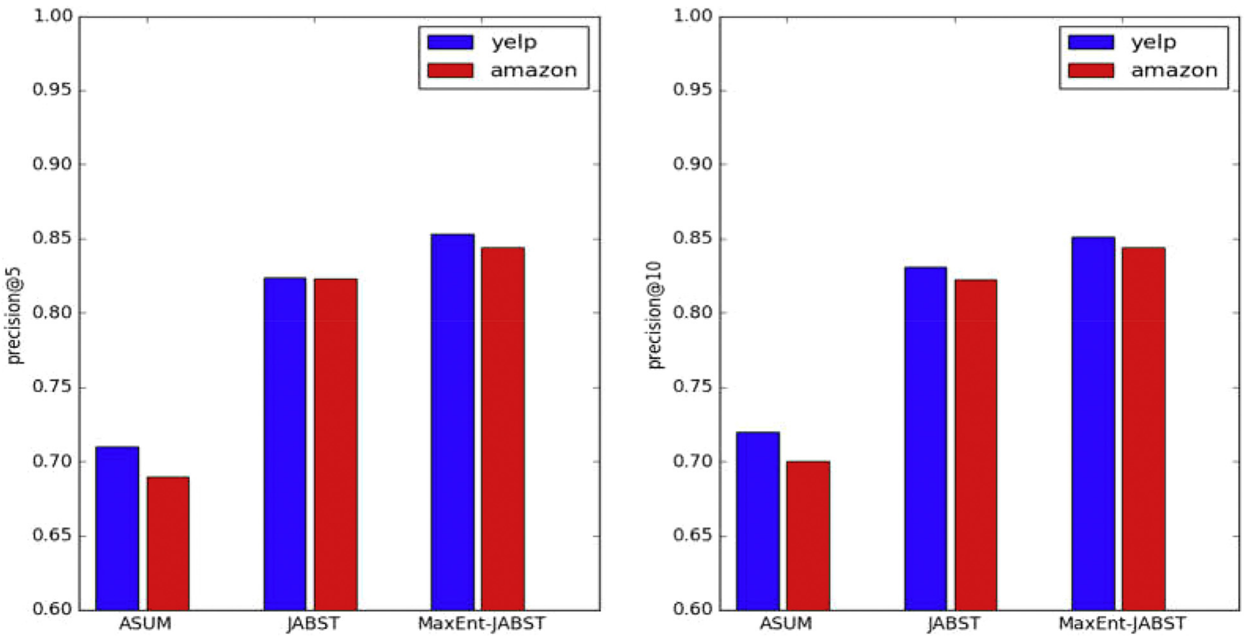

Fig. 6. Average precision on specific opinion identification.

Table 8
Examples of specific aspects identified from ASUM, JABST and MaxEnt–JABST.

Topic 1			Topic 2		
MaxEnt–JABST	JABST	ASUM	MaxEnt–JABST	JABST	ASUM
vacumm	power	vacuum	pizza	chees	pizza
hair	vacuum	hose	chees	pizza	crust
pick	cleaner	attach	sausag	crust	dish
filter	carpet	cord	sauc	*italian*	*place*
dyson	attach	*use*	tomato	pasta	chicago
suction	cord	power	oliv	salad	*one*
vac	*head*	handl	slice	mozza	pie
rug	floor	*get*	pasta	bread	*order*
attach	*tool*	*one*	mozzarella	oil	chees
wood	electrolux	*wand*	*flavor*	pie	*style*

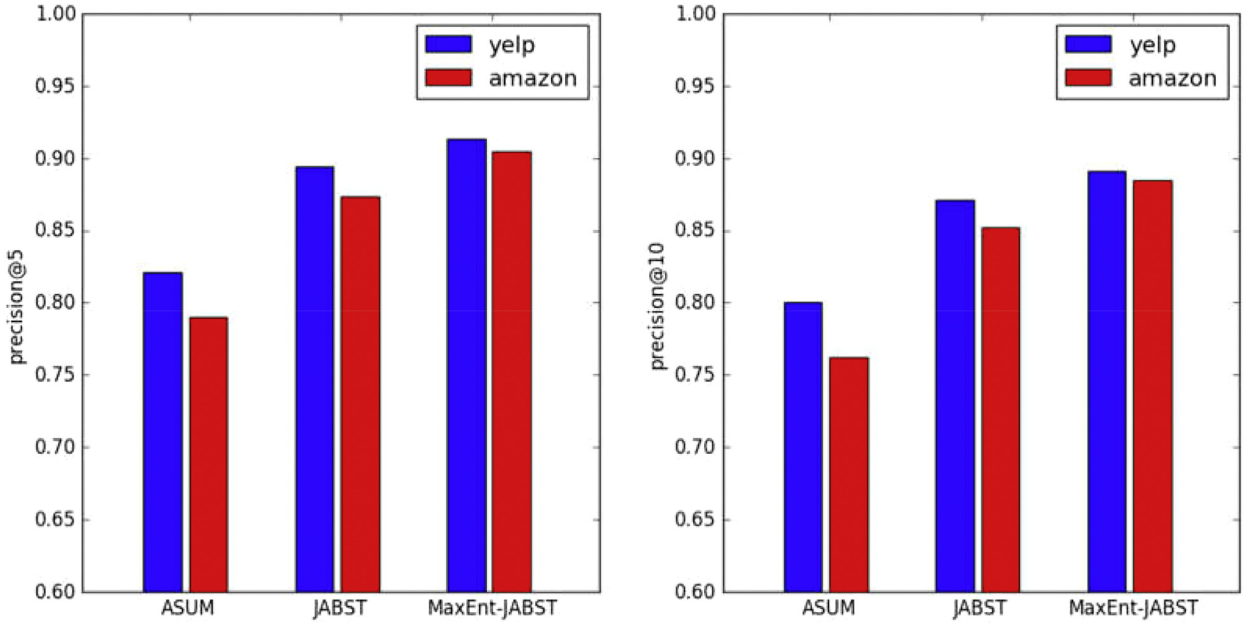

Fig. 7. Average precision on specific aspect identification.

the maximum entropy classifier, and corresponding classifiers are called as "*MaxEnt–JABST* + *A*" and "*MaxEnt–JABST* + *B*", respectively. Accordingly, our MaxEnt–JABST model uses input vectors with both the lexical features and the part of speech features.

We tested the specific opinion precision and specific aspect precision under the three different feature inputs. As shown Figs. 8 and 9, "MaxEnt–JABST + *B*" with only part of speech feature vectors has better precision than "MaxEnt–JABST + *A*" with only lexical feature vectors. This result means that part of speech features are more beneficial to feature selection than lexical features. Moreover, our MaxEnt–JABST model that uses both features has highest precision because lexical features are also helpful to feature selection.

5.3.6. Domain adaptability

To evaluate the adaptability to different domain datasets, we use the movie review[6] [19] as training dataset. Then, we tested the specific opinion precision and the specific aspect precision of our MaxEnt–JABST model. Tables 9 and 10 demonstrate that our MaxEnt–JABST can be used in various domains with different features. The reason is that our MaxEnt–JABST model uses the maximum entropy classifier and supervised learning, and considers both the lexical features and the part of speech features. This is another advantage of our MaxEnt–JABST model.

[6] Polarity dataset v2.0: https://www.cs.cornell.edu/people/pabo/movie-review-data/.

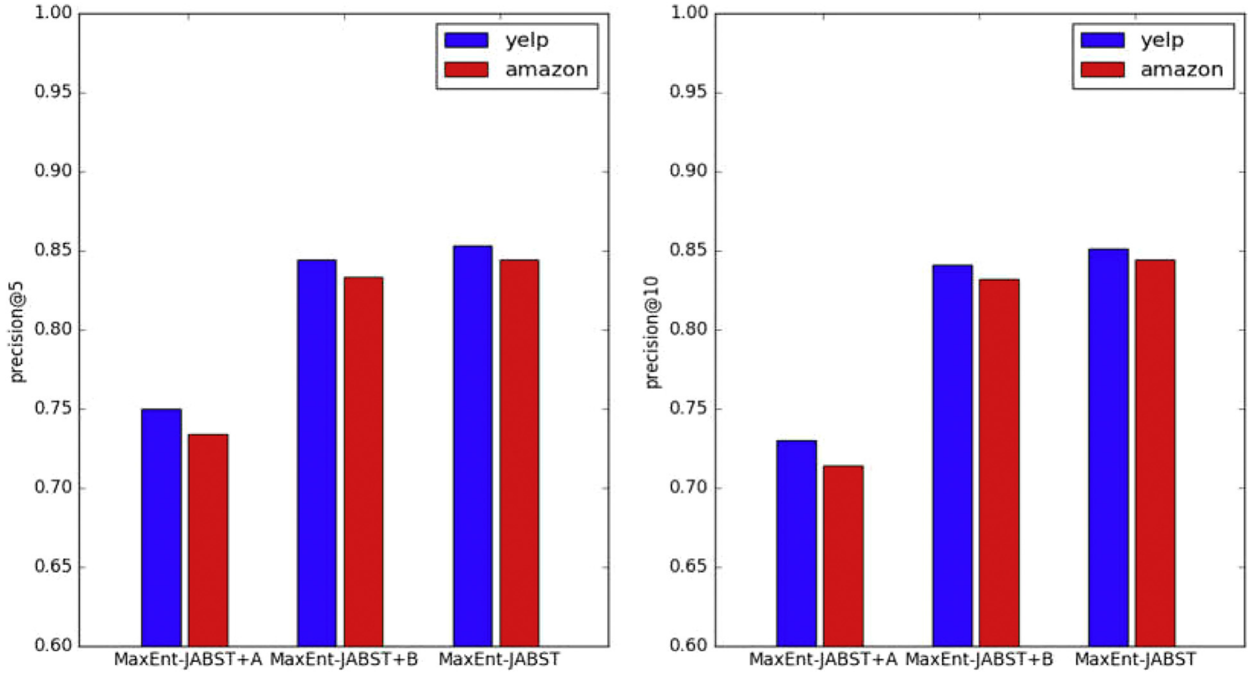

Fig. 8. Average specific opinion precision with different input feature vectors.

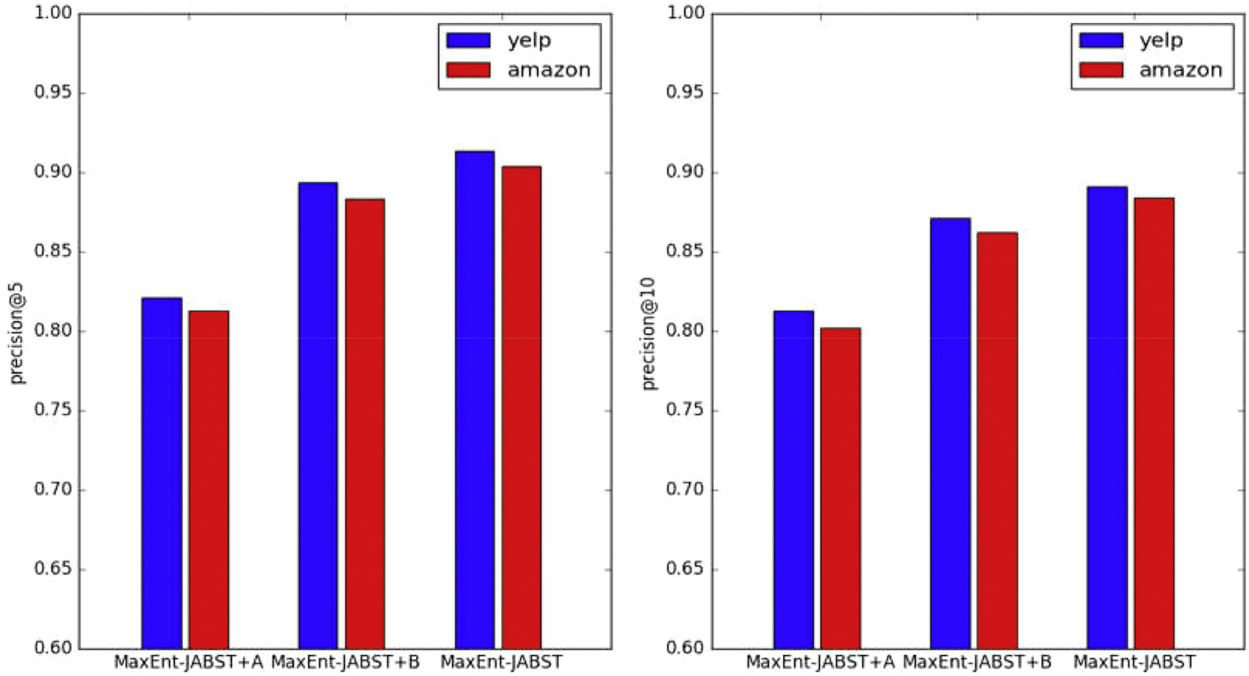

Fig. 9. Average specific aspect precision with different input feature vectors.

Table 9
Opinion precision of MaxEnt–JABST model with different domain datasets.

Datasets	Yelp	Amazon
precision@5	0.846	0.831
precision@10	0.829	0.816

Table 10
Aspect precision of MaxEnt–JABST model with different domain datasets.

Datasets	Yelp	Amazon
precision@5	0.892	0.887
precision@10	0.853	0.848

6. Conclusions

This paper presents the two generative models that can extract fine-grained aspects and opinions from online reviews. The proposed JABST model extracts general and specific aspects and opinions jointly, with their sentiment polarity. Furthermore, the MaxEnt–JABST adds a maximum entropy classifier to separate opinion and aspect words more accurately. Consequently, our approach can simultaneously handle all tasks in aspect-based sentiment analysis, including identifying multi-granularity aspects and opinions, and sentiment polarity. The evaluation results on the two real-world review datasets demonstrate that our models significantly outperform baseline models. Most importantly, our models are able to identify meaningful opinion words strongly associated with corresponding aspects, and specific aspect words associated with different aspects of entities.

Along with this direction, we are going to build hierarchical models [6] for aspects and opinions. Moreover, cross-domain or cross-linguistic sentiment analysis will be useful for more complex mobile application [2,23–25], which will be an extension of our work in the future.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China projects under Grants 61832013, 61672351, 61872235 and 61729202, and in part by the Huawei Technologies Co., Ltd project under Grant YBN2018125107.

References

[1] D.M. Blei, B.B. Edu, A.Y. Ng, A.S. Edu, M.I. Jordan, J.B. Edu, Latent Dirichlet allocation, J. Mach. Learn. Res. 3 (2003) 993–1022.
[2] Q. Diao, M. Qiu, C.-Y. Wu, A.J. Smola, J. Jiang, C. Wang, Jointly modeling aspects, ratings and sentiments for movie recommendation (JMARS), in: Proc. of SIGKDD, 2014, pp. 193–202.
[3] T. Hofmann, Learning the similarity of documents: an information-geometric approach to document retrieval and categorization, in: Advances in Neural Information Processing System, 2000, pp. 914–920.
[4] M. Hu, B. Liu, Mining and summarizing customer reviews, in: Proc. 2004 ACM SIGKDD Int. Conf. Knowl. Discov. Data Min. KDD 04, 2004.
[5] Y. Jo, Aspect and sentiment unification model for online review analysis, in: Proc. Fourth ACM Int. Conf. Web Search Data Min., 2011, pp. 815–824.
[6] S. Kim, J. Zhang, Z. Chen, A. Oh, S. Liu, A hierarchical aspect- sentiment model for online reviews, in: Proc. Twenty-Seventh AAAI Conf. Artif. Intell., 2013, pp. 526–533.
[7] H. Lakkaraju, C. Bhattacharyya, I. Bhattacharya, S. Merugu, Exploiting coherence for the simultaneous discovery of latent facets and associated sentiments, in: Proc. 2011 SIAM Int. Conf. Data Min., 2011, pp. 498–509.
[8] C. Lin, Y. He, R. Everson, S. Ruger, Weakly supervised joint sentiment-topic detection from text, IEEE Trans. Knowl. Data Eng. 24 (6) (2012) 1134–1145.
[9] Y. C.Lin, He. Joint sentiment/topic model for sentiment analysis, in: Proc. 18th ACM Int. Conf. Inf. Knowl. Manag. - CIKM '09, 2009, pp. 375–384.
[10] B. Liu, Sentiment analysis and opinion mining, Synth. Lect. Hum Lang. Technol. 5 (1) (2012) 1–167.
[11] D. Marcheggiani, O. Tackstrom, A. Esuli, F. Sebastiani, Hierarchical Multi-Label Conditional Random Fields for Aspect-Oriented Opinion Mining, in: Hierarchical Multi-Label Conditional Random Fields for Aspect-Oriented Opinion Mining, 2014, pp. 273–285.
[12] Q. Mei, X. Ling, M. Wondra, H. Su, C. Zhai, Topic sentiment mixture: modeling facets and opinions in weblogs, in: Proc. 16th Int. Conf. World Wide Web - WWW '07, 2007, pp. 171–180.
[13] D. Mimno, H.M. Wallach, E. Talley, M. Leenders, A. McCallum, Optimizing semantic coherence in topic models, in: Proc. 2011 Conf. Empir. Methods Nat. Lang. Process., 2, 2011, pp. 262–272.
[14] S. Moghaddam, M. Ester, ILDA: interdependent LDA model for learning latent aspects and their ratings from online product reviews categories and subject descriptors, in: Proc. 34th Int. ACM SIGIR Conf. Res. Dev. Inf. Retr., 2011, pp. 665–674.
[15] S. Moghaddam, M. Ester, The FLDA model for aspect-based opinion mining: addressing the cold start problem, in: Proc. of 22nd Int. Conf. World Wide Web (WWW 2013), 2013, pp. 909–918.
[16] A. Mukherjee, B. Liu, Aspect extraction through semi-supervised modeling, in: Proc. 50th Annu. Meet. Assoc. Comput. Linguist. Long Pap., 1, 2012, pp. 339–348.
[17] A. Mukherjee, B. Liu, Modeling review comments, in: Proc. of ACL-2012, 2012, pp. 320–329.
[18] D. Newman, Y. Noh, E. Talley, S. Karimi, T. Baldwin, Evaluating topic models for digital libraries, in: Proc. 10th ACM/IEEE-CS Jt. Conf. Digit. Libr., 2010, pp. 215–224.
[19] B. Pang, L. Lee, A sentimental education: sentiment analysis using subjectivity summarization based on minimum cuts, in: Proc. of the 42nd Annual Meeting on Association for Computational Linguistics, 2004.
[20] C.H. Papadimitriou, H. Tamaki, P. Raghavan, et al., Latent semantic indexing: a probabilistic analysis, in: Proc. of the Seventeenth ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems, ACM, 1998, pp. 159–168.
[21] K. Schouten, F. Frasincar, Survey on aspect-level sentiment analysis, IEEE Trans. Knowl. Data Eng. 28 (3) (2016) 813–830.
[22] K. Stevens, P. Kegelmeyer, D. Andrzejewski, D. Buttler, Exploring topic coherence over many models and many topics, in: Proc. 2012 Jt. Conf. Empir. Methods Nat. Lang. Process. Comput. Nat. Lang. Learn., July, 2012, pp. 952–961.
[23] F. Tang, M. Guo, S. Guo, C. Xu, Mobility prediction based joint stable routing and channel assignment for mobile ad hoc cognitive networks, IEEE Trans. Parallel Distrib. Syst. 27 (3) (2016) 789–802.
[24] F. Tang, J. Li, Joint rate adaptation, channel assignment and routing to maximize social welfare in multi-hop cognitive radio networks, IEEE Trans. Wirel. Commun. 16 (4) (2017) 2097–2110.
[25] F. Tang, H. Zhang, L.T. Yang, Multipath cooperative routing with efficient acknowledgement for LEO satellite networks, IEEE Trans. Mob. Comput. 18 (1) (2019) 179–192.

[26] I. Titov, R. McDonald, Modeling online reviews with multi-grain topic models, in: Proceeding 17th Int. Conf. World Wide Web, 2008, p. 111.
[27] I. Titov, R. McDonald, A joint model of text and aspect ratings for sentiment summarization, in: Proc. ACL08 HLT, 2008, pp. 308–316.
[28] H. Wang, Y. Lu, C. Zhai, Latent aspect rating analysis on review text data, in: Proc. 16th ACM SIGKDD Int. Conf. Knowl. Discov. Data Min. - KDD '10, 2010, p. 783.
[29] W.X. Zhao, J. Jiang, H. Yan, X. Li, Jointly modeling aspects and opinions with a MaxEnt-LDA hybrid, Comput. Linguist. 16 (2010) 56–65.
[30] L. Zhuang, F. Jing, X.-Y. Zhu, Movie review mining and summarization, in: Proc. 15th ACM Int. Conf. Inf. Knowl. Manag. - CIKM'06, 2006.