Aspect-Level Opinion Mining of Online Customer Reviews

XU Xueke, CHENG Xueqi, TAN Songbo, LIU Yue, SHEN Huawei

Key Laboratory of Web Data Science and Technology, Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100190, China

Abstract: This paper focuses on how to improve aspect-level opinion mining for online customer reviews. We first propose a novel generative topic model, the Joint Aspect/Sentiment (JAS) model, to jointly extract aspects and aspect-dependent sentiment lexicons from online customer reviews. An aspect-dependent sentiment lexicon refers to the aspect-specific opinion words along with their aspect-aware sentiment polarities with respect to a specific aspect. We then apply the extracted aspect-dependent sentiment lexicons to a series of aspect-level opinion mining tasks, including implicit aspect identification, aspect-based extractive opinion summarization, and aspect-level sentiment classification. Experimental results demonstrate the effectiveness of the JAS model in learning aspect- dependent sentiment lexicons and the practical values of the extracted lexicons when applied to these practical tasks.

Key words: online customer reviews; aspect-level opinion mining; aspect-dependent sentiment lexicon; Joint Aspect/Sentiment model

I. INTRODUCTION

With the emergence of Web 2.0, customers can freely write reviews about different entities, such as digital products or hotels, via various Web 2.0 platforms. Automatic opinion mining techniques extract, analyze and summarize the opinions [1-4] in a large number of reviews and thus help users quickly digest the opinions of interest. Since people tend to be more interested in particular aspects (for example, ambience or service of a restaurant) than the whole entity, opinion mining techniques aiming at different aspects rather than the whole entity are especially appealing and have gained much attention in recent years. Aspect-level opinion mining could help users effectively navigate into detailed information of their interesting aspects by organizing the opinion summarization in a structured form [1, 4-5].

To perform aspect-level opinion mining tasks, we need to find the major aspects of entities in a specific domain (e.g. restaurants). Furthermore, a high-quality sentiment lexicon plays a fundamental role in these tasks. However, a general-purpose sentiment lexicon is usually not favorable due to the highly aspect-dependent nature of sentiment [6-9]. Therefore, going beyond only finding aspects, we should further extract aspect-dependent sentiment lexicon for each major aspect, i.e. aspect-specific opinion words along with their aspect-aware sentiment polarities. This kind of lexicon would be potentially useful for improving aspect-level opinion mining task performance, which is embodied in the following three points.

Firstly, aspect-specific opinion words could help infer the targeted aspects in case that the aspects are not explicitly given such as in “so delicious!”, so called implicit aspect identification. Implicit aspect identification is a challenging and important problem [3], and we

Received: 2012-12-15
Revised: 2013-02-18
Editor: YUAN Baozong

This paper focuses on how to improve aspect-level opinion mining for online customer reviews. We propose the Joint Aspect/Sentiment model (JAS) to extract aspects and aspect-dependent sentiment lexicons from online customer reviews in a unified framework. The extracted sentiment lexicons are applied to a series of aspect-level opinion mining tasks, including implicit aspect identification, aspect-based extractive opinion summarization, and aspect-level sentiment classification.

argue that aspect-specific opinion words could provide rich clues for inferring the implicit aspects. Indeed, the customers tend to use opinion words specific to the aspect to comment on the aspect, e.g. using "cozy" and "romantic" to comment on the ambience of a restaurant. And these specific words could thus, in turn, help infer the targeted aspects.

Secondly, aspect-specific opinion words could also help extract more informative opinions from reviews. These opinion words usually provide more meaningful descriptions about the aspect [7]. They could tell users why or from what perspective the opinions about the aspect are favorable or unfavorable rather than giving only general information such as "good" or "bad" as the general opinion words.

Lastly, the knowledge of aspect-aware sentiment polarities could help aspect-level sentiment classification [8]. Indeed, the sentiment polarities of many opinion words are sensitive to the targeted aspects, which means that one single word may deliver different sentiment polarities according to the aspects in context or deliver a sentiment only for a specific aspect. For instance, for a hotel we enjoy a "large" room, but not expect "large" noise; it is desirable for the ambience of a restaurant to be "private", although "private" is generally neutral, etc. Identifying aspect-aware polarities for these opinion words could thus improve aspect-level sentiment classification, and consequently, help provide more accurate positive vs. negative statistic summary about the customers' opinions on the aspect.

Although general-purpose sentiment lexicon and domain-specific sentiment lexicon extraction has been well studied, aspect-dependent sentiment lexicon extraction is still a pilot task with less attention. As pioneering work, Samuel Brody in Ref. [10] and Yue Lu in Ref. [8] took a three-stage approach where aspect extraction, opinion words extraction and sentiment polarities determination is conducted separately. Largely different from their work, we took a unified approach to jointly extracting aspects and aspect-dependent sentiment lexicons. Specifically, we proposed a novel generative topic model, namely Joint Aspect/Sentiment (JAS) model, to extract aspects and aspect-dependent sentiment lexicons from online customer reviews in a unified framework. JAS extends Latent Dirichlet Allocation (LDA) model [11] in several ways in order to jointly address the following progressive challenges: 1) Detect major aspects of entities in the specific domain (e.g. restaurants); 2) Detect aspect-specific opinion words for each discovered aspect; 3) Identify aspect-aware sentiment polarities for the opinion words with respect to each aspect. Note that, several joint models involving aspect (or topic) and sentiment have also been developed by incorporating sentiment factor into the classic generative topic models. However, these models either could not further identify sentiment polarities for the extracted opinion words [7] or were not designed to explicitly model sentiment specific to the aspects [6, 12-13].We will give detailed comparisons between our model and two most related models in Section 4.3.

We applied the extracted sentiment lexicons to several aspect-level opinion mining tasks, including implicit aspect identification [3], aspect-based extractive opinion summarization [1, 4-5], and aspect-level sentiment classification [8]. More specifically, we used the knowledge of aspect-specific opinion words to identify implicit aspects and to find aspect-relevant and informative opinions in reviews; we used the aspect-aware sentiment polarities knowledge to help determine the sentiment of the opinions about the specific aspect in texts. Experimental results showed the practical values of the lexicons in helping these tasks.

II. Related Research Work

Aspect-dependent sentiment lexicon extraction is still a pilot task with less attention. As pioneering work, Samuel Brody in Ref. [10] and Yue Lu in Ref. [8] took a three-stage approach where aspect extraction, opinion words extraction and sentiment polarities determination is conducted separately. Brody and Elhadad in Ref. [10] first detected aspects using

Local LDA model, then selected adjectives from aspect-relevant texts as aspect-specific opinion words, and at last identified aspect-sensitive polarities of the adjectives using polarity propagation based on an aspect-specific polarity graph. Yue Lu in Ref. [8] proposed an optimization framework to combine different signals for determining aspect-aware sentiment polarities. The aspects were predefined aspects with manually selected keywords, and the opinion words were extracted beforehand using NLP techniques. Besides, that approach relies heavily on manually provided information, e.g. sentiment rating for each review, which is often unavailable in practice.

Several joint models involving aspect (or topic) and sentiment have been developed by incorporating sentiment factor into classic generative topic models (e.g., LDA [11], and pLSA [14]). However, these models were not designed to explicitly extract aspect-dependent sentiment-lexicon.

Topic Sentiment Mixture (TSM) model [13] was the first such joint model by integrating sentiment into pLSA. However the detected sentiments are general for all topics, while our model can detect aspect-specific sentiments.

Joint Sentiment/Topic (JST) model [6] was the first LDA based model to simultaneously consider topics and sentiments. JST does not aim to detect topic-specific sentiments, but rather detect sentiment-topic pairs, or sentiment-bearing topics under different sentiment labels [15], which help review-level sentiment classification.

Aspect and Sentiment Unification Model (ASUM) [12] follows a similar generative process to JST except that a sentiment-topic (aspect) pair is selected for a single sentence, rather than for a word as JST, such that the detected sentiment-topic pairs by ASUM fits the aspects of entities. ASUM, in essence, aims to detect sentiment-coupled aspects with respect to different sentiments rather than explicitly detecting sentiments specific to the aspects as our model.

MaxEnt-LDA [7] was the first to jointly discover both aspects and aspect-specific opinion words by integrating supervised maximum entropy (MaxEnt) component to separate opinion word from factual words. However, it does not further identify aspect-aware sentiment polarities, which is very important but challenging. Furthermore, MaxEnt-LDA uses some labeled data to learn the MaxEnt component.

There are also many joint models of aspects and sentiment ratings [16-18] which is, however, not the focus of this work. In this paper, "sentiment" refers to opinion words with their sentiment polarities, not numeric ratings.

III. THE JOINT ASPECT/SENTIMENT MODEL

JAS is a novel generative topic model that aims to extract aspects and aspect-dependent sentiment lexicons from online reviews in a given domain.

Firstly, we adapt the classic topic model LDA to make the extracted topics correspond to the reviewable aspects, rather than global properties, of entities by constraining that all words of each sentence are assigned to a single topic. The underlying observation is that each sentence tends to present a single aspect [12].

Secondly, we introduce two kinds of indicator variables, i.e. subjectivity label and sentiment label into the model in order to explicitly model the sentiment specific to the detected aspects. Specifically, for each aspect, our model could learn three multinomial distributions over words, which respectively model the factual semantics of the aspect, and the positive and negative sentiment specific to the aspect. Based on these word distributions, we could naturally construct the aspect-dependent sentiment lexicon for the specific aspect.

Compared to the previous work [8], our model needs no domain-specific knowledge resource or manually labeled data, which makes it highly portable across domains.

3.1 The generative process

Assume we have a corpus of D customer reviews in a specific domain (e.g. restaurant reviews), each review is a list of sentences, each sentence is a list of words, and each word is an entity from a vocabulary with V distinct words denoted by $w = 1, 2, \ldots, V$. Each sentence s in review d is associated with one variable: the aspect $z_{d,s}$ which is shared by all words in the sentence. And the nth word $w_{d,s,n}$ in the sentence s of review d is associated with two indicator variables: the subjectivity label $\zeta_{d,s,n}$ and the sentiment label $l_{d,s,n}$. Here, $\zeta_{d,s,n}$ indicates whether $w_{d,s,n}$ is a sentiment-conveying opinion word ($\zeta_{d,s,n} = \mathbf{opn}$) or a factual word (i.e. not conveying sentiment) ($\zeta_{d,s,n} = \mathbf{fact}$). And $l_{d,s,n}$ indicates whether $w_{d,s,n}$ conveys a positive sentiment ($l_{d,s,n} = \mathbf{pos}$) or a negative sentiment ($l_{d,s,n} = \mathbf{neg}$). We now give an intuitive description of how a review is generated according to our model.

For each sentence s in the review d, we draw an aspect $z_{d,s}$ from a distribution over T aspects conditioned on d. Then, the following steps will be taken to generate each word $w_{d,s,n}$ in the sentence s:

1) We draw a subjectivity label $\zeta_{d,s,n}$ from a distribution over subjectivity labels {**opn**, **fact**}, $\upsilon^{d,s,n}$, to indicate that whether $w_{d,s,n}$ is sentiment-conveying or factual.

2) We then draw a sentiment label $l_{d,s,n}$ from a distribution over sentiment labels {**pos**, **neg**} conditioned on the subjectivity label (either **opn** or **fact**, as indicated by $\zeta_{d,s,n}$) and the sentence s.

3) If $\zeta_{d,s,n} = \mathbf{opn}$, the word $w_{d,s,n}$ will be generated from a distribution over words conditioned on the sentiment (either positive or negative, as indicated by $l_{d,s,n}$) that is specific to the aspect $z_{d,s}$. If $\zeta_{d,s,n} = \mathbf{fact}$, $w_{d,s,n}$ will be generated from a distribution over words conditioned on the factual aspect $z_{d,s}$.

Note that, the distribution over subjectivity labels, i.e. $\upsilon^{d,s,n}$, is key for appropriate subjectivity label assignment for $w_{d,s,n}$, we will discuss in details how to set this distribution in Section 3.2. Also note that, when $\zeta_{d,s,n} = \mathbf{fact}$, $l_{d,s,n}$ is meaningless since $w_{d,s,n}$ does not convey any sentiment in that case, and will actually be ignored for generating $w_{d,s,n}$. We here draw it just for completeness of the generation process, and not drawing it is also reasonable.

The formal generative process is as follows:

1. For each aspect t, draw a multinomial distribution over words: $\Phi^t \sim Dir(\beta)$
   (a) For each sentiment specific to the aspect t, draw a multinomial distribution over words, respectively:
   $\Phi^{t,\mathrm{pos}} \sim Dir(\beta^{\mathrm{pos}})$, $\Phi^{t,\mathrm{neg}} \sim Dir(\beta^{\mathrm{neg}})$
2. For each review d in the corpus:
   (a) Draw a multinomial distribution over aspects $\theta^d \sim Dir(\alpha)$
   (b) For each sentence s in review d:
      (i) Draw an aspect $z_{d,s} \sim \theta^d$
      (ii) For each subjectivity label $\zeta \in$ {**fact**, **opn**}, draw a Bernoulli distribution over sentiment labels $\pi^{d,s,\zeta} \sim Beta(\gamma)$
      (iii) For each word $w_{d,s,n}$ in the sentence s:
         (1) Choose a subjectivity label $\zeta_{d,s,n} \sim \nu^{d,s,n}$
         (2) If $\zeta_{d,s,n} = \mathbf{opn}$:
            (a) Choose a sentiment label $l_{d,s,n} \sim \pi^{d,s,\mathrm{opn}}$
            (b) Generate the word $w_{d,s,n} \sim \Phi^{z_{d,s}, l_{d,s,n}}$
         (3) If $\zeta_{d,s,n} = \mathbf{fact}$:
            (a) Choose a sentiment label $l_{d,s,n} \sim \pi^{d,s,\mathrm{fact}}$
            (b) Generate the word $w_{d,s,n} \sim \Phi^{z_{d,s}}$

Figure 1 shows the graphical representation of the generation process, where $\boldsymbol{S} = 2$ and $\boldsymbol{J} = 2$ are the numbers of sentiment labels and subjectivity labels respectively, $\boldsymbol{M}_d$ is the number of sentences in the review d, and $\boldsymbol{N}_{d,s}$ is the number of words in the sentence s of the review d.

3.2 Separating opinion words from factual words

Appropriate subjectivity label assignments for words in reviews is a key for detecting aspect-specific opinion words. And the subjectivity label distribution $\nu^{d,s,n}$ plays an important role in subjectivity label assignment for $w_{d,s,n}$. However, fully-unsupervised topic models, which mainly exploit co-occurrences of words to detect latent topics, cannot effectively separate opinion words from factual words since these two kinds of words are usually mixed together in texts. Therefore, inspired by the work in Ref. [7], instead of drawing $\nu^{d,s,n}$ from a symmetric Beta prior, we could set $\nu^{d,s,n}$ by applying various external sources of knowledge (presented by λ in Figure 1) to the context features of the word $w_{d,s,n}$ (presented by $c_{d,s,n}$ in Figure 1) to indicate the probability of whether or not $w_{d,s,n}$ conveys a sentiment.

In the current instantiation of JAS, we consider only the word itself as its context feature, and integrate the knowledge from an opinion word lexicon for setting $\nu^{d,s,n}$. The knowledge of this lexicon is encoded into the parameters $\{\lambda^w | w \in \{1,2,...,V\}\}$, where λ^w is a distribution over subjectivity labels for the word w, and we have $\lambda_{\text{opn}}^w + \lambda_{\text{fact}}^w = 1$. Specifically, for each word w in the opinion lexicon, we set λ_{opn}^w to a value approaching 1, e.g. 0.95 as in our experiments; while for each word w not contained by the lexicon, we set λ_{opn}^w to a value approaching 0, e.g. 0.05 as in our experiments. Then, based on the opinion word lexicon knowledge, $\nu^{d,s,n}$ could be set as follows:

$$P(\zeta_{d,s,n} = \zeta \mid w_{d,s,n}) = \nu_\zeta^{d,s,n} = \frac{\lambda_\zeta^{w_{d,s,n}}}{\lambda_{\text{opn}}^{w_{d,s,n}} + \lambda_{\text{fact}}^{w_{d,s,n}}}$$

$$\zeta \in \{\mathbf{opn}, \mathbf{fact}\}$$

In this way, the subjectivity label assignment for $w_{d,s,n}$ is, to a large degree, decided by whether the word $w_{d,s,n}$ is contained by the lexicon. If contained, $w_{d,s,n}$ will tend to be assigned to subjectively label **opn**, if not, it will tend to be assigned to **fact.**

It is worth noting that, our model is very flexible to incorporate more sources of knowledge and more context features of $w_{d,s,n}$ to better identify sentiment-bearing words.

3.3 Model inference & sentiment lexicon extraction

First we give explanations for the notations used in this section in Table I.

In order to estimate the word distributions for the factual aspect (i.e. Φ^t), and the aspect-specific positive and negative sentiments (i.e. $\Phi^{t,\text{pos}}$ and $\Phi^{t,\text{neg}}$), we first use the collapsed

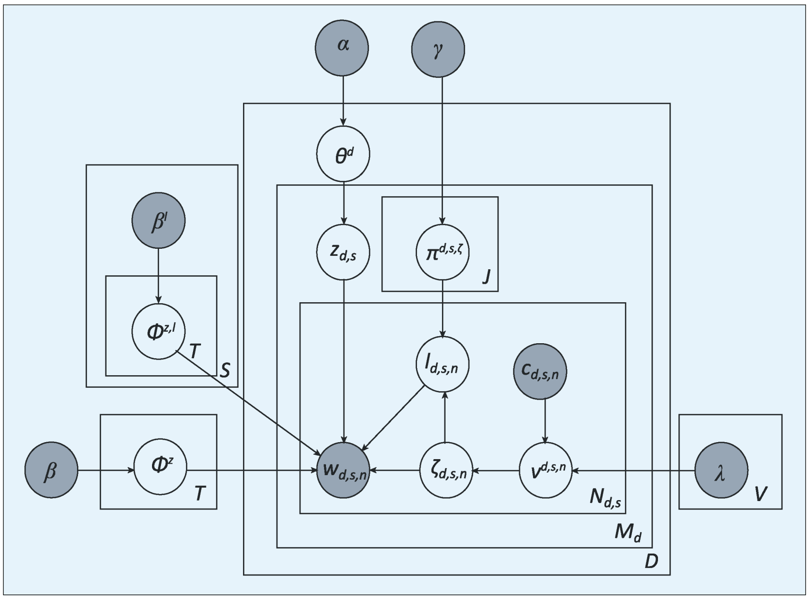

Fig.1 *The graphical representation of JAS*

Table I *The explanations of notations*

$\boldsymbol{w}$	The word list of the entire corpus.
z ($z_{\neg(d,s)}$)	The aspect assignments over the list of sentences (excluding sentence s in review d).
ζ ($\zeta_{\neg(d,s,n)}$)	The subjectivity label assignments over the list of words (excluding the nth word in the sentence s of review d,$w_{d,s,n}$).
$\boldsymbol{l}$ ($\boldsymbol{l}_{\neg(d,s,n)}$)	The sentiment label assignments over the list of words (excluding the nth word in the sentence s of review d, $w_{d,s,n}$).
c^d (c_t^d)	The number of sentences (assigned to aspect t) in review d.
$n^{d,s,\text{fact}}$ ($n_w^{d,s,\text{fact}}$)	The number of times any word (or word w) assigned to subjectivity label **fact** in sentence s of review d.
$n^{d,s,\text{opn},l}$ ($n_w^{d,s,\text{opn},l}$)	The number of times any word (or word w) assigned to subjectivity label **opn** and sentiment label l in sentence s of review d. $l \in \{$**pos**, **neg**$\}$
$c^{t,\text{fact}}$ ($c_w^{t,\text{fact}}$)	The number of times any word (or word w) assigned to subjectivity label **fact** under aspect t.
$c_l^{d,s,\text{opn}}$	The number of words in sentence s assigned to subjectivity label **opn** and sentiment label l.
$c^{t,\text{opn},l}$ ($c_w^{t,\text{opn},l}$)	The number of times any word (or word w) assigned to subjectivity label **opn** and sentiment label l under aspect t.

Gibbs sampling [19] to estimate the posterior distributions over $\boldsymbol{z}$, $\boldsymbol{\zeta}$ and $\boldsymbol{l}$. According to the collapsed Gibbs sampling, each variable of interest (e.g. $z_{d,s}$) will be sequentially drawn according to a probability distribution conditioned on current assignments for all other variables and the observed data.

Specifically, we first draw $z_{d,s}$ by the following conditional probability:

$$P(z_{d,s} = t \mid \boldsymbol{z}_{\neg(d,s)}, \boldsymbol{\zeta}, \boldsymbol{l}, \boldsymbol{w}) \propto \frac{c_t^d + \alpha_t}{\sum_{t'=1}^{T} (c_{t'}^d + \alpha_{t'})} \times$$

$$\left(\left(\frac{\Gamma\left(\sum_{w=1}^{V}(c_w^{t,\text{fact}}+\beta_w)\right)}{\Gamma\sum_{w=1}^{V}(c_w^{t,\text{fact}}+\beta_w)+n^{d,s,\text{fact}})}\right)\cdot\right.$$

$$\left.\prod_{w=1}^{V}\left(\frac{\Gamma(c_w^{t,\text{fact}}+\beta_w+n_w^{d,s,\text{fact}})}{\Gamma(c_w^{t,\text{fact}}+\beta_w)}\right)\right)\times$$

$$\prod_{l\in\{\text{pos},\text{neg}\}}\left(\left(\frac{\Gamma\left(\sum_{w=1}^{V}(c_w^{t,\text{opn},l}+\beta_w^l)\right)}{\Gamma\left(\sum_{w=1}^{V}(c_w^{t,\text{opn},l}+\beta_w^l)+n^{d,s,\text{opn},l}\right)}\right)\cdot\right.$$

$$\left.\prod_{w=1}^{V}\left(\frac{\Gamma(c_w^{t,\text{opn},l}+\beta_w^l+n_w^{d,s,\text{opn},l})}{\Gamma(c_w^{t,\text{opn},l}+\beta_w^l)}\right)\right) \quad (1)$$

Here Γ is the gamma function. Note that all the numbers represented by c in above the equation exclude sentence s of review d.

Then, we could jointly draw values for $\zeta_{d,s,n}$ and $l_{d,s,n}$ as the following conditional probabilities:

$$P(\zeta_{d,s,n}=\textbf{fact}\mid \boldsymbol{z},\boldsymbol{\zeta}_{\neg(d,s,n)},\boldsymbol{l}_{\neg(d,s,n)},\boldsymbol{w})$$
$$\propto\frac{\lambda_{\text{fact}}^{w_{d,s,n}}}{\sum_{\zeta\in\{\textbf{opn},\textbf{fact}\}}\lambda_{\zeta}^{w_{d,s,n}}}\times\frac{c_{w_{d,s,n}}^{t,\text{fact}}+\beta_{w_{d,s,n}}}{\sum_{w=1}^{V}(c_w^{t,\text{fact}}+\beta_w)} \quad (2)$$

$$P(\zeta_{d,s,n}=\textbf{opn},l_{d,s,n}$$
$$=l\mid \boldsymbol{z},\boldsymbol{\zeta}_{\neg(d,s,n)},\boldsymbol{l}_{\neg(d,s,n)},\boldsymbol{w})$$
$$\propto\frac{\lambda_{\text{opn}}^{w_{d,s,n}}}{\sum_{\zeta\in\{\textbf{fact},\textbf{opn}\}}\lambda_{\zeta}^{w_{d,s,n}}}\times$$
$$\frac{c_l^{d,s,\text{opn}}+\gamma_l}{\sum_{l'\in\{\textbf{pos},\textbf{neg}\}}(c_{l'}^{d,s,\text{opn}}+\gamma_{l'})}\times$$
$$\frac{c_{w_{d,s,n}}^{t,\text{opn},l}+\beta_{w_{d,s,n}}^l}{\sum_{w=1}^{V}(c_w^{t,\text{opn},l}+\beta_w^l)};l\in\{\textbf{neg},\textbf{pos}\} \quad (3)$$

Note that, all the numbers represented by c in the above two equations exclude the word $w_{d,s,n}$.

The detailed Gibbs sampling procedure of the JAS is shown in follows.

Initialization of assignments for all variables.
For n = 1 to N Gibbs sampling iterations.
1. For each sentence s (suppose from the review d)
   a) Exclude the sentence s from the numbers represented by c involved in Eq. (1).
   b) Calculate the conditional probability of assigning sentence s to aspects as Eq.(1).
   c) Choose an aspect assignment for the sentence s based on the computed probability.
   d) Include the sentence s into the numbers with the new aspect assignment.
   e) For each word $w_{d,s,n}$ in the sentence s:
      i. Exclude the word from the numbers represented by c involved in Eq. (2) and Eq. (3).
      ii. Calculate the conditional probabilities for assignments of subjectivity label and sentiment label for the word as Eq. (2) and Eq. (3).
      iii. Jointly choose subjectivity label and/or sentiment label assignments for the word based on the computed probabilities.
      iv. Include the word into the numbers with the new assignment results.

Based on the last sample of $\boldsymbol{z}$, $\boldsymbol{\zeta}$ and $\boldsymbol{l}$, the word distributions for the factual aspect (i.e. Φ^t) and the aspect-specific positive and negative sentiments (i.e. $\Phi^{t,\text{pos}}$ and $\Phi^{t,\text{neg}}$) can be approximated as follows:

$$\Phi_w^t=\frac{c_w^{t,\text{fact}}+\beta_w}{c^{t,\text{fact}}+\sum_{w'=1}^{V}\beta_{w'}}$$

$$\Phi_w^{t,l}=\frac{c_w^{t,\text{opn},l}+\beta_w^l}{c^{t,\text{opn},l}+\sum_{w'=1}^{V}\beta_{w'}^l};l\in\{\textbf{neg},\textbf{pos}\}$$

Then we have a set of extracted aspects. For each aspect t, top probability words in Φ^t present its factual semantic information; top probability words in $\Phi^{t,\text{pos}}$ and $\Phi^{t,\text{neg}}$ are aspect-specific opinion words from, and their sentiment polarities respecting the aspect could be determined as follows: if $\Phi_w^{t,\text{pos}}>\Phi_w^{t,\text{neg}}$, the word w is positive for the aspect t; otherwise negative.

3.4 Incorporating sentiment prior

Sentiment Prior (SP) knowledge serves as guidance for identifying sentiment polarities of the opinion words. Here, sentiment prior means a set of SP words (usually a subset of an opinion word lexicon) along with their prior sentiment labels. We here have two parts of SP words: Soft SP words and Hard SP words. A Hard SP word, such as “excellent”, will convey the same sentiment as the prior in any context. A Soft SP word will deliver the sentiment as the prior in most contexts, but with exceptions.

We incorporate the sentiment prior into our model by using asymmetric β^{pos} and β^{neg} which give Dirichlet priors of $\Phi^{t,\text{pos}}$ and $\Phi^{t,\text{neg}}$ respectively. These two priors describe our assumptions of the word distributions for the positive and negative sentiments for any aspect before observing the data. Specifically, for each positive Hard SP word w, β_w^{neg} is set to 0. Similarly, for each negative Hard SP word w, β_w^{pos} is set to 0. Besides, in the initialization step of the Gibbs sampling, all Hard SP words are assigned to their prior sentiment labels. In this way, we could impose the hard-constraints that the Hard SP words could only be assigned to their prior sentiment labels in the Gibbs sampling process. For each positive Soft SP word, β_w^{neg} is set to a relatively smaller value compared with β_w^{pos}. Similarly, for each negative Soft SP word, β_w^{pos} is set to a relatively smaller value compared with β_w^{neg}. In this way, we impose the soft-constraints that the Soft SP words are more probable to be assigned to their prior sentiment labels. Note that, the soft-constraints could be relaxed, i.e. the sentiment labels of these words would be adjusted in the Gibbs Sampling process. For all other words in the vocabulary, both β_w^{pos} and β_w^{neg} are set to the same value, which means we have no prior assumption on the sentiment labels of these words.

Intuitively speaking, our model propagates the sentiment prior information, via aspect-contextual sentence-level co-occurrences of opinion words in reviews, in a bootstrapping-like manner to adapt and extend sentiment prior with respect to the aspect. The underlying observation is that a single sentence tends to present one sentiment, either positive or negative [12], and thus opinion words tend to convey the same sentiment with the prior sentiment label of co-occurring SP words in the sentence.

IV. The Evaluation of the Model

4.1 Experimental settings

1) Data

To evaluate our model, we used the two freely available data sets, the restaurant reviews and the hotel reviews initially used in Ref. [20] and Ref. [21], respectively. The restaurant reviews have been preprocessed with sentence segmentation and part-of-speech tagging. For hotel reviews, we used a NLP toolkit[1] to segment reviews into sentences, and used the Stanford POS Tagger[2] to conduct part-of-speech tagging over sentences. Since negating words, such as “not”, usually change the polarity of the opinion word, we added a negation prefix, “not_”, to a word modified by a negating word. We then removed stop words based on a stop word list[3]. Finally, each sentence was converted to a list of POS tagged words with possible negation prefix. For instance, the sentence “the quality is not good” would be finally converted into “quality_noun not_good_adj”.

2) Opinion word lexicon

The opinion word lexicon (see Section 3.2) came from two widely used knowledge bases: MPQA Subjectivity Lexicon[4] (MPQA in short) and SentiWordNet[5]. For the SentiWordNet part, we selected a list of words with a positive or negative score above a given threshold. As for the MPQA part, we extracted another list of words with “type” being “strongsubj” or “weaksubj”. Finally we constructed the lexicon as union of the two parts. Note that, the words in both MPQA and SentiWordNet are all POS tagged words, such as “good_adj”.

3) Sentiment Prior

The SP words (see Section 3.4) are actually a subset of the opinion word lexicon. To develop sentiment prior, we selected as SP words only the words from MPQA part with words

[1] http://l2r.cs.uiuc.edu/~cogcomp/atool.php?tkey=SS
[2] http://nlp.stanford.edu/software/tagger.shtml
[3] http://ir.dcs.gla.ac.uk/resources/linguistic_utils/stop_words/
[4] http://www.cs.pitt.edu/mpqa/
[5] http://sentiwordnet.isti.cnr.it/

of “neutral priorpolarity” filtered out, and used their “priorpolarity” in MPQA as their prior sentiment labels. We used MPQA as the source of sentiment prior knowledge because MPQA could provide high precision sentiment polarities [8], while the sentiment polarities in SentiWordNet were inferred automatically and not reliable enough.

From SP words, we further selected words with all senses (i.e. synsets in SentiWordNet) sharing the same sentiment polarity according to the SentiWordNet. We then manually checked these selected words, trying to ensure that they deliver consistent sentiment in any context, and finally obtained the Hard SP words. The remaining SP words were Soft SP words.

Note that, the development of sentiment prior was totally domain-independent and aspect-independent.

4) Parameters settings

We ran 200 iterations of Gibbs sampling, which was adequate for obtaining good and stable results in our experiments. We set $\alpha_t = 50/T$ for each aspect t and $\beta_w = 0.1$ for each word w in the vocabulary as in Ref. [19]. The default value for β_w^{pos} and β_w^{neg} was set to 0.1 as β_w. But for each positive Hard SP word w, β_w^{neg} was set to 0, for each negative Hard SP word w, β_w^{pos} was set to 0, for each positive Soft SP word w, β_w^{neg} was set to 0.01, and for each negative Soft SP word w, β_w^{pos} was set to 0.01. Besides, γ_l was set to 0.001 for each sentiment label l, and $\{\lambda^w | w \in \{1,2,...,V\}\}$ were set as in Section 3.2. The aspect number T was set to 10 with which we could detect all major aspects, and no more meaningful aspects has been detected by increasing T.

4.2 Sample results

Tables II and III show sample results with restaurant and hotel reviews, respectively. For space limitation, we here only show some major aspects, and discard some miscellaneous (e.g. “Anecdote”) or similar (e.g. “Food-Main dish”) aspects. The tables clearly show that our model could effectively extract both aspects and aspect-dependent sentiment lexicon knowledge.

1) Factual aspects

Generally, our model could effectively detect major aspects from both restaurants and hotel reviews. And the extracted factual words are quite coherent and meaningful with respect to the aspects. For instance, with restaurants reviews, our model could detect major aspects of restaurants (i.e. “Service”, “Food”, and “Ambience”), and further find more fine-grained aspects, such as “Service-Order Taking” and “Service-Staff”. For hotel reviews, our model could also effectively discover major aspects of hotels such as “Service-Reception”, “Room Condition”, “Location” and “Breakfast”.

2) Aspect-specific opinion words

The discovered opinion words (either positive or negative) are quite specific and informative with respect to the aspects. For instance, the staff is “knowledgeable”, the cream is “sour”, the location of the hotel is “central”, etc. Our model could also well capture the differences in opinion word usages even for related aspects in different domains. Take the food related aspects as examples. In restaurant reviews, we mainly see some quite specific opinion words, such as “greasy” and “fresh”, focusing on the food quality. In hotel reviews, the opinion words are more comprehensive, discussing not only quality of food at breakfast but also other issues related to breakfast, such as availability. These observations conform to the practical differences in customers’ concerns about the food aspect of a restaurant and a hotel.

3) Aspect-aware sentiment polarities

In Tables II and III, some interesting words are marked with * to make them notable. These words either show strongly aspect-dependent sentiment polarities (e.g. “small”, “young” and “private”) or are not included in the SP words (e.g. “casual”, “greasy”, and “unprofessional”). The tables show that, generally speaking, our model could effectively identify sentiment polarities for opinion words with respect to the aspects. For instance, in restaurant reviews, “long” waiting time is not acceptable for reservation, it is desirable for the ambience to be

Table II *Sample results with restaurant reviews*

Food-Bakery	**Factual**	chocolate cheese dessert cake cream bread ice fries tea pizza pie eggs sauce desserts
	Neg.	sour hard cold tasted frozen bad tiny insult burnt* weak soggy* # plain dry* stale greasy* bland* small* heavy*
	Pos.	best delicious fresh good hot sweet great crust amazing perfect # especially tasty* excellent huge*
Service-Staff	**Factual**	staff service attentive_noun wait waiter waiters waitstaff owner waitress servers server atmosphere customers manager
	Neg.	slow rude poor attitude* bad horrible inattentive obnoxious terrible unprofessional* # problem unfriendly intrusive arrogant incompetent clueless* unapologetic*
	Pos.	friendly attentive nice helpful extremely prompt great professional* courteous knowledgeable # polite warm fast*
Service-Order Taking	**Factual**	table minutes seated wait reservation told waiter hour manager waited away order hostess
	Neg.	rude little busy extremely cold long* completely* actually* hungry unfriendly # unprofessional* late* arrogant unapologetic*
	Pos.	friendly promptly good great immediately* sure nice happy worth ready # attentive large helpful professional*
Ambience	**Factual**	room music bar atmosphere decor tables dining space place area seating crowd restaurant walls garden ambience
	Neg.	tiny loud* uncomfortable dim complaint hard ergonomic difficult obnoxious stuffy # tacky* fake dim
	Pos.	nice beautiful romantic great cozy small cool* fun* comfortable warm live # open intimate casual* private* elegant modern

*The aspect names are manually given for demonstration. To improve readability we remove the POS tags from words unless needed. For "**Factual**", we show exactly top words. While for "**Neg.** " and "**Pos.** " words before the separator "#" are top 10 words, and the followings are selected words from top 100 to supplement the results. Some words are marked with * to make them notable. Note that, we only consider those words with $\Phi_w^{t,\text{pos}} > \Phi_w^{t,\text{neg}}$ ($\Phi_w^{t,\text{pos}} <= \Phi_w^{t,\text{neg}}$) for sample positive (negative) opinion words.

Table III *Sample results with hotel reviews*

Service-Reception	**Factual**	staff english desk spoke reception service help directions speak concierge recommendations questions
	Neg.	trouble rude problem unfriendly poor bad arrogant hard miserable unfortunately # complaint unprofessional* young* loud*
	Pos.	helpful friendly extremely excellent great good pleasant clean nice polite # courteous efficient attentive professional*
Location	**Factual**	location walk walk_noun distance steps station walking minutes metro spanish minute trevi vatican
	Neg.	problem hard bad unsafe seedy slow chaos difficult affordable hassle # busy inconvenient
	Pos.	easy great close* major* quiet* good short* convenient central* perfect # safe easily conveniently nearby* near*
Food-Breakfast	**Factual**	breakfast coffee buffet fruit eggs cheese juice included cereal continental tea selection pastries bread cereals served
	Neg.	hard limited poor disappointing terrible stale complaint bad mediocre horrible # cheap hungry problem sour bland
	Pos.	good fresh cold hot nice great delicious excellent plentiful adequate typical* # plenty tasty* available*
Room Condition	**Factual**	room bathroom rooms shower bed size beds bath decorated bathrooms towels double marble tv water tub
	Neg.	tiny small* hard problem uncomfortable single* dirty old* dark complaint # cold common* narrow smelly* nasty stained*
	Pos.	clean comfortable large nice spacious modern good nicely big* huge* # new* beautiful spotless quiet* decent soft

*The explanations of this table are the same with Table II.

"private", etc. In hotel reviews, people prefer for "central" and "close" hotels, "small" room is not desirable, the waiter is "young" means he/she is not experienced, etc. However, we also observe some incorrect cases, such as "cheap" for the location of hotels. In the future, we plan to incorporate more sources of signals, such as "and" rules in linguistics heuristics and synonym/antonym rules [8], to better identify aspect-aware sentiment polarities.

4.3 Comparisons with related models

In this section, we will compare our model with the state-of-art joint models of aspect and sentiment, i.e. MaxEnt-LDA [7] and ASUM [12], with the help of some result samples.

1) MaxEnt-LDA

To the best of our knowledge, MaxEnt-LDA is the first to jointly extract aspects and aspect-specific opinion words. However, it does not further identify aspect-aware sentiment polarities. Besides, MaxEnt-LDA uses some labeled data to learn the syntactic patterns, while our approach needs no labeled data.

We here compared our model with MaxEnt-LDA in terms of how well they could detect aspect-specific opinion words. Seen from Tables II and III in this paper and Tables III, V in

Table IV *A comparison sample for "Order Taking" aspect with restaurant reviews*

(a) MaxEnt-LDA	
Factual	table minutes wait waiter reservation order time hour manager people
Opinion	seated asked told waited waiting long arrived rude sat finally
(b) JAS	
Factual	table minutes seated wait reservation told waiter hour manager waited
Neg.	rude little busy extremely cold long completely actually hungry unfriendly
Pos.	friendly promptly good great immediately sure nice happy worth ready

Table V *A sample results with restaurant reviews by ASUM*

Negative Aspect2	table waiter manager told time minutes said people order restaurant party hostess waitress check food service bar left staff
Positive Aspect2	good best great food pizza cheese place fresh delicious fries make better coffee chicken burger hot really chocolate bread

Ref. [7], we could observe that, in general, our model could detect more specific and informative opinion words. We also asked two human assessors to manually judge whether the top 10 positive and negative opinion words detected by JAS and top 10 opinion words by MaxEnt-LDA were really opinion words with clear association with the aspects[6]. Averaging over judgments by the two assessors, only 61.4% by MaxEnt-LDA are clearly aspect-related opinion words, compared to 77.8% by our model. Note that the results by MaxEnt-LDA in Ref. [7] are based on the same restaurant and hotel review data sets as ours.

We presented a sample result with "Order Taking" aspect in Table IV to give a further comparison. MaxEnt-LDA mainly leverages syntactic patterns, encoded into a maximum entropy component, to separate opinion words from factual words. Consequently, some factual words that often play an opinion expressing role in sentences according to the syntactic patterns (e.g., "asked" in "the waiter is asked") will be incorrectly identified as opinion words. We could see such words as "seated", "asked", and "told" incorrectly identified as opinion words for "Order Taking" of restaurants in Table V (a). Comparatively, our approach could avoid such pseudo opinion words by integrating a high-quality opinion word lexicon.

2) ASUM

ASUM aims to detect sentiment-coupled aspects with respect to different sentiments rather than explicitly detecting sentiments specific to the aspects as our model. The differences between ASUM and our model could be clearly illustrated by the results in Ref. [12] as well as the results shown in Table V. The results in Table V were obtained by running ASUM over the restaurant reviews based on the implementation by the authors of ASUM[7] with some minor improvements to incorporate more sentiment prior knowledge. The aspect number was set to 10, and the other parameters in AUSM were set as in Ref. [12].

In the generation process of ASUM, for each sentence, the sentiment label is first drawn, and the aspect is then drawn conditioned the sentiment label. All words in the sentence will be finally generated from the word distribution for the sentiment-coupled aspect without distinguishing between sentiment-bearing opinion words and factual words. Thus, we could only learn sentiment-fact mixed knowledge rather than pure sentiment lexicon knowledge. Indeed, as shown in Tables V and VI in Ref. [12], most sentiment- coupled aspects learned by ASUM are actually dominated by factual words.

To detect aspect-specific opinion words, additional steps should be taken (see Section 6.3 of Ref. [12]). However, in Table VII of Ref. [12], we still observe that the opinion words and the factual words are highly mixed. For instance, the learned top negative opinion words for aspect "Service" are "said me want card get tell if would gui bad could rude pai becaus walk then".

V. The Application of the Lexicons

We applied the aspect-dependent sentiment lexicons extracted by JAS to a series of practical aspect-level opinion mining tasks, including implicit aspect identification, aspect-based extractive opinion summarization, and aspect-level sentiment classification. We quantitatively tested the quality of the aspect-dependent sentiment lexicons according to how well they could improve the performance of these tasks. These evaluations could provide more objective and reliable judgments compared with di-

[6] For the restaurant reviews, we consider "Food-Bakery", "Service-Staff", "Service-Order Taking", and "Ambience"aspects in Table II, corresponding to "Food", "Staff", "Order-Taking", and "Ambience" aspects in Table III of Ref. [21] respectively; for hotel reviews, we consider "Food-Breakfast", "Service-Reception", "Room Condition" aspects in Table III, corresponding to "Meal", "Service", "Room Condition" aspects in Table V of Ref. [21], respectively.

[7] http://uilab.kaist.ac.kr/research/WSDM11

rect manual judgments, which may introduce personal biases, as well as highlight the utility of our model in practical applications.

5.1 Implicit aspect identification

The task of "implicit aspect identification" is to infer the targeted aspect by the opinion words when the factual aspect words are absent. This is a challenging but important task, since implicit aspects are common in online views [3]. We argue that aspect-specific opinion words could provide rich clues for inferring the implicit aspects. For instance, in sentence "so delicious!" aspect words are not explicitly given, but we still could easily infer the underlying aspect, i.e. "Food", according to the opinion word "delicious".

We quantitatively tested how well JAS could extract aspect-specific opinion words by applying this knowledge to "implicit aspect identification". The evaluations were based on a set of approximately 3 400 sentences manually annotated with aspect and sentiment polarity information by Ganu et al. [20] in the restaurant reviews (see Section 4.1). We selected from the gold standard aspect set (i.e. {"Staff", "Food", "Ambience", "Price", "Anecdote", "Misc"}) three major aspects: "Staff", "Food", and "Ambience". We discarded the other aspects because of the same reasons as in Ref. [7]: 1) "Price" is often mixed with other aspects such as "Food", and 2) "Anecdote" and "Misc" do not show clear semantics. Refs. [7, 10] also only used these three aspects for evaluations of aspect identification with the restaurant reviews. Note that, we made the factual aspect words in the sentences implicit by using only opinion words for aspect identification.

To use this gold standard data for evaluations, we manually found automatically detected aspects that correspond to each gold standard aspect. Since JAS could detect quite fine-grained aspects, there may be multiple detected aspects, such as "Food-Bakery" and "Food-Main dish", for one gold standard aspect, such as "Food". For each gold standard aspect a, we could rank all the sentences in the gold standard data according to the negative KL-divergence between the Aspect-specific Opinion Model (AOM) and the sentence language model. Then we could use Precision at different rank positions to measure the performance. AOM reflects the aspect-specific opinion words, and could be learned by combining both positive and negative opinion words as follow:

$$P(w \mid \theta_a^{\text{opn}}) \propto \begin{cases} \dfrac{\sum\limits_{l\in\{\text{neg,pos}\}} \left(\sum\limits_{t\in T_a} c_w^{t,\text{opn},l} + \beta_w^l \right)}{\sum\limits_{l\in\{\text{neg,pos}\}} \left(\sum\limits_{t\in T_a} c^{t,\text{opn},l} + \sum\limits_{w'=1}^{V} \beta_{w'}^l \right)}; w \in V_O \\ 0\,; w \notin V_O \end{cases}$$

where T_a is the detected aspects by JAS that corresponds to a, and the vocabulary V_O is exactly the opinion word lexicon. We here only retained probability values for opinion words such that we could purely use opinion words for aspect identification. The sentence model was estimated using maximum likelyhood estimation with Dirichlet smoothing.

We considered for comparisons following alternative approaches to learning the AOM.

Gen: This approach takes AOM as a general opinion model with each opinion word uniformly distributed.

Bo1: This approach can be summarized as follows: First, we pick up the sentences annotated with the given aspect, called Aspect Sentences (AS). Then, we use the Bo1 model [22] to assign a weight to each word in the opinion word lexicon, measuring how discriminative it is in AS against the whole restaurant reviews, to infer the probability of the word in the AOM.

Note that, because the training data required by MaxEnt-LDA [7] was not available, we here could not use MaxEnt-LDA to learn AOM.

Seen from Figure 2, the performance of our approach is remarkably better than **Gen**, and is close to **Bo1** over three aspects. When N is small (e.g. $<=200$), our approach is even comparable to, if not better than, **Bo1** with very high precision. Note that, **Bo1** is a fully

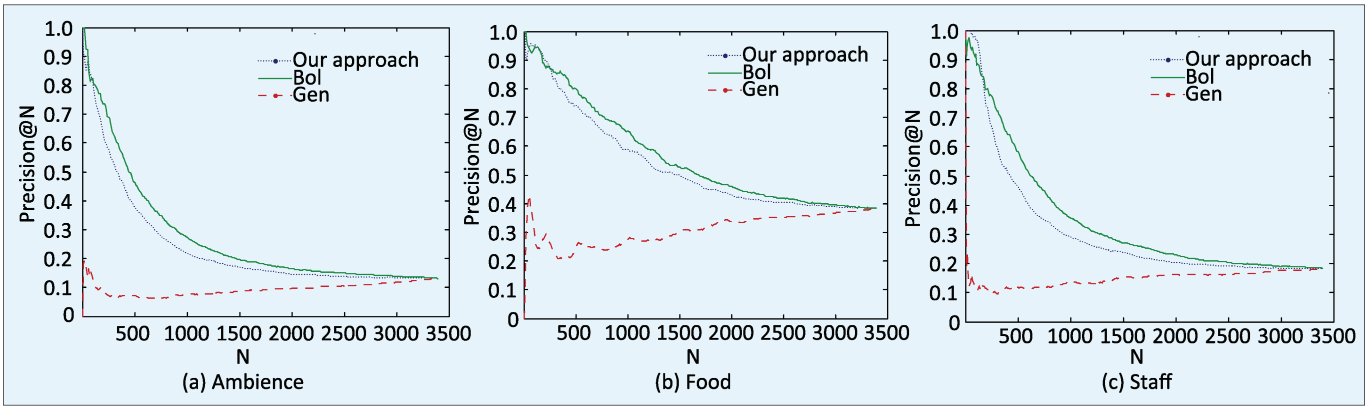

Fig.2 *Precision @ N curves with the three aspects. Here N is a position in the sentence rank.*

supervised strong approach. It could even be considered as an approximate upper bound for an unsupervised approach like ours, since it is directly trained based on manually annotated data and the Bo1 word weighting model which is effective in discriminating specific words from those general [22]. Even further, **Bo1** takes advantages on that the testing data is the same with the training data.

These observations show our model could effectively detect opinion words that are quite specific to the aspects, which could help identify aspect-relevant opinions from reviews when the factual aspect words are not explicitly given.

5.2 Aspect-based extractive opinion summarization

Given a set of reviews on a specific entity (e.g. a restaurant), for each of major aspects, this task is to extract a small number of sentences with aspect-relevant opinions to deliver the major opinions about the aspect in the reviews [1, 4, 9]. We tested how well the aspect-specific opinion words extracted by JAS could help extract relevant and informative opinions when applied to this task.

We conducted evaluations based on the restaurant reviews. For evaluation, we selected 10 restaurants just with most review numbers, and use the same three major aspects (i.e. "Food", "Staff", and "Ambience") as in the Section 5.1. We evaluated the summary quality using ROUGE toolkit, which was officially adopted by TAC for automatic summarization evaluation. ROUGE criteria measure summary quality by counting overlapping units such as the n-grams between the generated summary and the reference summary (i.e. how well the generated summary fits the reference summary). For each aspect of a restaurant, we created the reference summary by picking several sentences (about 100 words in total) with relevant, informative, and non-redundant opinions, excluding those with only general opinions (e.g. "great" and "bad") or redundant opinions. One such reference summary is shown in Table VI.

The summarization approach is straightforward. We rank sentences according to the negative Kullback-Leibler (KL) divergence between the sentence language model and a mixture model of the Factual Aspect Model (FAM) and the Aspect-specific Opinion Model (AOM). We could then generate the summary by picking sentences from the top ranked sentences until a given summary length limitation (i.e. 100 words) is reached. Note that, our approach is quite similar to that in Ref. [23], and the major difference is that we used a fact-opinion mixture model to improve the summary quality while Ref. [23] only used FAM.

For the aspect a, the mixture model could be defined as:

$$p(w\,|\,\theta_a^{\text{mix}}) = \lambda \cdot p(w\,|\,\theta_a^{\text{opn}}) + (1-\lambda) \cdot p(w\,|\,\theta_a^{\text{fact}})$$

where the FAM could be learned as:

$$p(w\,|\,\theta_a^{\text{fact}}) \propto \frac{\sum_{t \in T_a} c_w^{t,\text{fact}} + \beta_w}{\sum_{t \in T_a} c^{t,\text{fact}} + \sum_{w=1}^{V} \beta_w}$$

Table VI *Sample summary results for the "Staff" of a restaurant*

Reference Summary	1. Service is mediocre at best and even food has gone downhill. 2. truly awful service and unimpressive food 3. Service was attentive and friendly. 4. The services was extremely slow, or food when it finally got in out table was cold! 5. The wait staff and sommelier were knowledgeable and helpful, but I would only go back during the off hours or when the hype dies down. 6. The wait staff was friendly and attentive. 7. Service was quite good, as our waiter was knowledgeable about the ingredients we were unfamiliar with.
Extracted Summary ($\lambda = 0$)	1. Waited 18 minutes before anyone offered us water! Then another 15 minutes before anyone came for our dinner order. 2. When the busser finally cleared them, we waited another 20 minutes for the waiter to bring our check. 3. The service is attentive, yet our waiter was like a robot. I got the impression that servers are not allowed to speak to the customers at all, and answers about dishes were explained in a text-book sort of way. 4. I went'opening night' and after waiting 40 minutes, the hostess had messed up her wait list due to a convoluted (but cute) Italian city table system where they have contrived.
Extracted Summary ($\lambda = 1$)	1. It is great for celebrations; the reservation staff is friendly, and the table service is always knowledgeable and courteous. 2. Service was attentive and friendly. 3. I thought the pizza was wonderful, the waitstaff extremely friendly and the wait was a little long, but what do you expect ...get there early and stop complaining! 4. The wine list is impressive with wines from every region in Italy and the sommeliers are incredibly knowledgeable and friendly. 5. Staff was very friendly and helpful and I loved the atmosphere—I am in my 20's so I like loud, trendy places for a sat night.

The AOM could be learned by JAS (denoted as **JAS**) or alternative approaches (e.g. **Bo1** and **Gen**), as in Section 5.1. The parameter λ indicates the relative importance of AOM. When $\lambda = 0$, the mixture model will degenerate into FAM, and when $\lambda = 1$ the mixture model is actually the AOM.

Generally speaking, the factual aspect-relevance is less guaranteed by AOM compared with FAM. However, AOM could help determine whether the sentences contain informative and aspect-relevant opinions, while FAM ignores whether the sentences contain opinions.

In Figure 3, we show ROUGE-1 Average-F score curves against λ values for different approaches to learning AOM. Note that, we use the "–l" option in ROUGE toolkits to truncate summaries longer than 100 words for fair comparisons. We could observe that the score curve for **JAS** shows a remarkable rise as increasing λ value until 0.5. This shows that AOM learned by JAS really help extract relevant and informative opinions. When the λ value approaches 1, the curve for **JAS** shows a decrease. It is because that overly emphasizing AOM may lead to extracting less aspect-relevant opinions compared with using FAM. Note that, when $\lambda = 1$, we observe that AOM itself

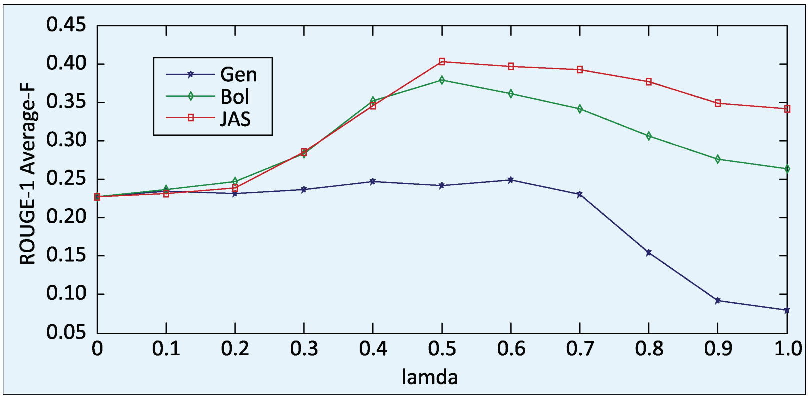

Fig.3 *ROUGE-1 Average-F score curves against λ values for different approaches to learning AOM*

could obtain considerable performance. Figure 3 also shows **JAS** consistently and remarkably outperforms **Gen** and the supervised approach **Bo1** when $\lambda \geqslant 0.5$, further showing the effectiveness of **JAS**.

Table VI gives sample summary results for the "Staff" of the restaurant with the most reviews. This table illustrates that the AOM learned by JAS could indeed help extract more informative opinions. When $\lambda = 0$, i.e. only FAM is used, we mainly see some factual descriptions about the "Staff" aspect; when $\lambda = 1$ we observe that even AOM itself could extract highly relevant as well as quite informative opinions.

5.3 Aspect-level sentiment classification

Aspect-level sentiment classification is to determine the sentiment polarity of opinions about a specific aspect in texts [8]. The evaluations were based on the labeled sentences in the restaurant reviews using the same three major aspects (i.e. "Staff", "Food", and "Ambience") as in Section 5.1. In order to avoid ambiguity, we only used sentences annotated with "positive" or "negative" polarity discarding those with "neutral" or "mix" polarity. Then, our specified task is to determine the sentiment polarity, either positive or negative, of opinions in the sentences annotated with the specific aspect.

Given a gold standard aspect a, our model could learn aspect-aware sentiment polarities for the opinion words used in Section 3.2. Specifically, we first learn positive and negative aspect-specific sentiment models:

$$p(w \mid \theta_a^{\text{opn},l}) \propto \frac{\sum_{t \in T_a} c_w^{t,\text{opn},l} + \beta_w^l}{\sum_{t \in T_a} c^{t,\text{opn},l} + \sum_{w'=1}^{V} \beta_w^l}; \quad l \in \{\mathbf{pos},\mathbf{neg}\}$$

Then the aspect-specific sentiment polarity of the opinion word w could be defined as: if $P(w \mid \theta_a^{\text{opn,pos}}) > P(w \mid \theta_a^{\text{opn,neg}})$, the word w is positive for the aspect a; otherwise it is negative.

Based on the aspect-aware sentiment polarities, our approach classifies the sentiment by just counting positive and negative opinion words in the sentences.

As comparisons, we have some general-purpose sentiment lexicon based baselines: **MPQA**, **SWN**, and **Union**, which use following sentiment lexicon respectively: *MPQA*, *SWN* and *Union*. *MPQA* contains words from the MPQA part in the opinion word lexicon (see Section 4.1), with "neutral" words filtered out, along with their "priorpolarity" in the MPQA. Actually, *MPQA* is exactly the sentiment prior used in our model (see Section 4.1). *SWN* contains words from the SentiWordNet part with their polarities inferred by SentiWordNet. *Union* is the union of *MPQA* and *SWN*. It actually contains all words in the opinion word lexicon, and we address possible conflicting polarities in *Union* by considering *MPQA* first.

We also adapted ASUM [12] to extracting aspect-dependent sentiment lexicon. The opinion words in the lexicon are the same as our approach, and the aspect-aware sentiment polarities are learned as follows. We first learn sentiment-coupled aspect models for aspect a as follows:

$$p(w \mid \theta_a^{AUSM,l}) \propto \frac{\sum_{t \in T_a^l} C_{ltw}^{STW} + \beta_{lw}}{\sum_{t \in T_a^l} \sum_{w'=1}^{V} C_{ltw'}^{STW} + \sum_{w'=1}^{V} \beta_{lw'}}; \quad l \in \{\mathbf{pos},\mathbf{neg}\}$$

where T_a^l is the detected aspects that corresponds to a under sentiment l, C_{ltw}^{STW} is the total number of times word w assigned to aspect t under sentiment l. Note that, since ASUM detects aspects under different sentiments independently, T_a^{pos} is usually different from T_a^{neg}. For an opinion word w, the sentiment polarity respecting the aspect a could be defined as: if

$$P(w \mid \theta_a^{AUSM,\text{pos}}) > P(w \mid \theta_a^{AUSM,\text{neg}}),$$

the word w is positive; otherwise negative. There are two variables for ASUM based approaches. **ASUM** incorporate the same sentiment prior, i.e. PARADIGM+ in Table III of Ref. [12], as the original ASUM, the aspect number is set to 10 as our approach, and the other parameters are the same as in [12]. **ASUM+** improves **ASUM** by using the same sentiment prior as our approach. We incorporate the sentiment prior by setting asymmetric β_l as in Section 3.4.

Besides lexicon-based unsupervised approaches, we also applied the state-of-the-art supervised learning approach, Support Vector Machine (**SVM**). Specifically, we used the LibSVM[8] to train the classifier based on the annotation information with all default options but a linear kernel[9]. Each sentence was represented by Vector Space Model with Term Frequency word weighting. The reported results for **SVM** were based on 7-fold cross validation.

[8] http://www.csie.ntu.edu.tw/~cjlin/libsvm/

[9] We have also used the default kernel, but it worked poorly.

[10] This paper is an extension to our previous conference short paper published in CIKM 2012[9].

Note that, Yue Lu in Ref. [8] also extracted aspect-dependent sentiment lexicon for aspect-level sentiment classification. However, their approach relies heavily on manually provided information, e.g., predefined aspects with manually selected keywords, sentiment rating for each review, etc, which were unavailable in the restaurant reviews.

Seen from Table VII, we could observe that our approach significantly outperforms the general-purpose sentiment lexicon based baselines. In particularly, we could see: 1) **MPQA** significantly outperforms both **SWN** and **Union.** This shows that *MPQA* could provide high precision polarities for opinion words in most contexts, while *SWN* provides low precision polarities and thus is not very reliable to be used as sentiment prior. 2) Our approach further outperforms **MPQA** over all three aspects. This is because that there are many opinion words, which deliver sentiment polarities strongly depending on the aspect or only carry sentiments for the specific aspect, not covered by *MPQA*. Our model effectively exploits aspect-contextual sentence-level co-occurrences of opinion words in reviews, to adapt and extend the knowledge of *MPQA* with respect to the aspect.

Our approach is even comparable to, if not better than, the state-of-art supervised learning approach **SVM** over the three aspects. Note that, our approach is extremely simple and efficient by only counting the numbers of positive and negative words, and needs no domain-specific manual labor, while **SVM** suffers from both high computational complexity and intensive labeling labors. Furthermore, we expect a promising potential of our model in improving aspect-level sentiment classification when combined with supervised learning approach.

We could also observe that, both **ASUM** and **ASUM+** perform poorly. We argue the reason is that, ASUM essentially aims to detect aspects conditioned on sentiment rather than detecting sentiment conditioned on aspects, and the sentiment-bearing opinion words and factual words are not explicitly separated. Thus, AUSM could not fully focus on aspect-contextual co-occurrences of sentiment-bearing words to learn sentiment polarities with respect to a specific aspect.

Table VII *Results of aspect-level sentiment classification in precision* (%)

	Food	Ambience	Staff	Avg.
JAS	82.83	81.59	78.19	80.87
MPQA	75.97	75.07	76.71	75.92
SWN	58.27	59.49	66.36	61.37
Union	70.50	67.42	72.09	70.00
SVM	83.99	80.15	79.30	81.15
ASUM	68.53	64.02	71.04	67.86
ASUM+	75.09	71.94	68.17	71.73

VI. Conclusion

In this paper, we attempted to extract aspects and aspect-dependent sentiment lexicon using a proposed Joint Aspect/Sentiment model. In future, we plan to improve our model in two folds. Firstly, we plan to consider more context information and more sources of knowledge to better identify opinion words. Secondly, we plan to incorporate more sources of signals, such as "and" rules in linguistics heuristics and synonym/antonym rules [8], to better identify aspect-aware sentiment polarities.

Acknowledgement

This work was supported by National Natural Science Foundation of China under Grants No. 61232010, No. 60903139, No. 60933005, No. 61202215, No.61100083; National 242 Project under Grant No. 2011F65; China Information Technology Security Evaluation Center Program under Grant No. Z1277.

References

[1] HU Minqing, LIU Bing. Mining and Summarizing Customer Reviews[C]// Proceedings of SIGKDD: August 22-25, 2004. Seattle, WA, USA. ACM Press, 2004: 168-177.

[2] GAMON M, AUE A, CORSTON-OLIVER S, *et al.* Pulse: Mining Customer Opinions from Free Text[C]// Proceedings of the 6th International Symposium on Intelligent Data Analysis: September 8-10, 2005. Madrid, Spain, 2005: 121-132.

[3] LIU Bing, HU Minqing, CHENG Junsheng. Opinion Observer: Analyzing and Comparing Opinions on the Web[C]// Proceedings of the 14th International Conference on World Wide Web: May 10-14, 2005. Chiba, Japan. ACM Press, 2005: 342-351.

[4] BLAIR-GOLDENSOHN S, NEYLON T, HANNAN K, *et al*. Building a Sentiment Summarizer for Local Service Reviews[C]// Proceedings of WWW Workshop on NLP in the Information Explosion Era: April 22, 2008. Beijing, China, 2008.

[5] XU Xueke, Meng Tao, CHENG Xueqi. Aspect-Based Extractive Summarization of Online Reviews[C]// Proceedings of the 2011 ACM Symposium on Applied Computing (SAC'11): March 21-24, 2011. Taichung, Taiwan, China. ACM Press, 2011: 968-975.

[6] LIN Chenghua, HE Yulan. Joint Sentiment/ Topic Model for Sentiment Analysis[C]// Proceedings of the 18th ACM Conference on Information and Knowledge Management: November 2-6, 2009. Hong Kong, China. ACM Press, 2009: 375-384.

[7] ZHAO W Xin, JIANG Jing, YAN Hongfei, *et al*. Jointly Modeling Aspects and Opinions with a MaxEnt-LDA Hybrid[C]// Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing: October 9-11, 2010. Cambridge, Massachusetts, USA, 2010: 56-65.

[8] LU Yue, CASTELLANOS M, DAYAL U, *et al*. Automatic Construction of a Context-Aware Sentiment Lexicon: an Optimization Approach[C]// Proceedings of the 20th International Conference on World Wide Web: March 28-April 1, 2011. Hyderabad, India. ACM Press, 2011: 347-356.

[9] XU Xueke, TAN Songbo, LIU Yue, *et al*. Towards Jointly Extracting Aspects and Aspect-Specific Sentiment Knowledge[C]// Proceedings of the 21st ACM International Conference on Information and Knowledge Management (CIKM'12): October 29-November 2, 2012. Maui, USA. ACM Press, 2012: 1895-1899.

[10] BRODY S, ELHADAD N. An Unsupervised Aspect-Sentiment Model for Online Reviews [C]// Proceedings of Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics: June 1-6, 2010. Los Angeles, California, USA, 2010: 804-812.

[11] BLEI D M, NG A Y, JORDAN M I. Latent Dirichlet Allocation[J]. Journal of Machine Learning Research, 2003, 3(3): 993-1022.

[12] JO Y, OH A H. Aspect and Sentiment Unification Model for Online Review Analysis[C]// Proceedings of the 4th ACM International Conference on Web Search and Data Mining: February 9-12, 2011. Hong Kong, China. ACM Press, 2011: 815-824.

[13] MEI Qiaozhu, LING Xu, WONDRA M, *et al*. Topic Sentiment Mixture: Modeling Facets and Opinions in Weblogs[C]// Proceedings of the 16th International Conference on World Wide Web: May 8-12, 2007. Banff, Alberta, Canada. ACM Press, 2007: 171-180.

[14] HOFMANN T. Probabilistic Latent Semantic Indexing[C] // Proceedings of the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval: August 15-19, 1999. Berkeley, California, USA. ACM Press, 1999: 50-57.

[15] HE Yulan, LIN Chenghua, ALANI H. Automatically Extracting Polarity-Bearing Topics for Cross-Domain Sentiment Classification[C]// Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: June 19-24, 2011. Portland, Oregon, USA, 2011: 123-131.

[16] TITOV I, MCDONALD R T. A Joint Model of Text and Aspect Ratings for Sentiment Summarization[C]// Proceedings of 46th Annual Meeting of the Association for Computational Linguistics: June 15-20, 2008. Columbus, Ohio, USA, 2008: 308-316.

[17] LAKKARAJU H, BHATTACHARYYA C, BHATTACHAYA I, *et al*. Exploiting Coherence in Reviews for Discovering Latent Facets and Associated Sentiments[C]// Proceedings of 2011 SIAM International Conference on Data Mining: April 28-30, 2011. Mesa, Arizona, USA, 2011.

[18] WANG Hongning, LU Yue, ZHAI Chengxiang. Latent Aspect Rating Analysis without Aspect Keyword Supervision[C]// Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining: August 21-24, 2011. San Diego, California, USA. ACM Press, 2011: 618-626.

[19] GRIFFITHS T L, STEYVERS M. Finding Scientific Topics[J]. Proceedings of the National Academy of Sciences, 2004, 101(Suppl 1): 5228-5535.

[20] GANU G, ELHADAD N, MARIAN A. Beyond the Stars: Improving Rating Predictions Using Review Text Content[C]// Proceedings of International Workshop on the Web and Databases: June 28, 2009. Providence, Rhode Island, USA, 2009.

[21] BACCIANELLA S, ESULI A, SEBASTIANI F. Multi-facet Rating of Product Reviews[C]// Proceed-

ings of the 31th European Conference on IR Research on Advances in Information Retrieval: April 6-9, 2009. Toulouse, France, 2009: 461-472.
[22] HE B, MACDONALD C, HE J, *et al*. An Effective Statistical Approach to Blog Post Opinion Retrieval[C]// Proceedings of the 17th ACM Conference on Information and Knowledge Management: October 26-30, 2008. Napa Valley, California, USA, 2008: 1063-1072.
[23] LING Xu, MEI Qiaozhu, ZHAI Chengxiang, *et al*. Mining Multi-faceted Overviews of Arbitrary Topics in a Text Collection[C]// Proceedings of the 14th ACMSIGKDD International Conference on Knowledge Discovery and Data Mining: August 24-27, 2008. Las Vegas, Nevada, USA. ACM Press, 2008: 497-505.

Biographies

XU Xueke, is currently a Ph.D. candidate at Institute of Computing Technology, Chinese Academy of Sciences, China. His research interests include web opinion retrieval and mining.

CHENG Xueqi, is a Professor at the Institute of Computing Technology, Chinese Academy of Sciences (ICT-CAS), China, and Director of Key Laboratory of Web Data Science and Technology in ICT-CAS. His main research interests include network science, Web search and data mining, P2P and distributed system, and information security. He has published over 100 papers in prestigious journals and international conferences, including New Journal of Physics, Physical Review E, Journal of Statistical Mechanics, IEEE Trans. Information Theory, SIGIR, WWW, CIKM, WSDM, IJCAI and so on. He is a recipient of the Youth Science and Technology Award of Mao Yisheng Science and Technology Award (2008), the CVIC Software Engineering Award (2008), the Excellent Teachers Award from Graduate University of Chinese Academy of Sciences (2006), the Second-Class Prize for the National Science and Technology Progress (2004), the Second-Class Prize for the Chinese Academy of Sciences and Technology Progress (2002), the Young Scholar Fellowship of Chinese Academy Sciences (2000).

TAN Songbo, received his Ph.D. degree in computer science from Institute of Computing Technology, Chinese Academy of Sciences, China in 2006. He is currently an Associate Professor in the Institute of Computing Technology, Chinese Academy of Sciences. His research interests include sentiment analysis, text classification, and machine learning.

LIU Yue, received her Ph.D. degree in computer science from Institute of Computing Technology, Chinese Academy of Sciences, China in 2004. She is currently an Associate Professor in the Institute of Computing Technology, Chinese Academy of Sciences. Her research interests include P2P and distributed system, and text mining.

SHEN Huawei, received the B.E. degree in electronic information from Xi'an Institute of Posts and Telecommunications, China, in 2003, and the Ph.D. degree in information security from the Institute of Computing Technology, the Chinese Academy of Sciences (CAS), China in 2010. Dr. Shen is currently an Associate Professor in Institute of Computing Technology, CAS. His major research interests include network science, information recommendation, user behaviour analysis, machine learning, and social network. He has published more than 20 papers in prestigious journals and top international conferences, including Physical Review E, Journal of Statistical Mechanics, Physica A, WWW, CIKM, and IJCAI. He is the member of the Association of Innovation Promotion for Youth of CAS. He was the recipient of the Top 100 Doctoral Thesis Award of CAS in 2011. He was awarded the Grand Scholarship of the President of CAS in 2010.