Analyzing Sentiments in One Go: A Supervised Joint Topic Modeling Approach

Zhen Hai, Gao Cong, Kuiyu Chang, Peng Cheng, and Chunyan Miao

Abstract—In this work, we focus on modeling user-generated review and overall rating pairs, and aim to identify semantic aspects and aspect-level sentiments from review data as well as to predict overall sentiments of reviews. We propose a novel probabilistic supervised joint aspect and sentiment model (SJASM) to deal with the problems in one go under a unified framework. SJASM represents each review document in the form of opinion pairs, and can simultaneously model aspect terms and corresponding opinion words of the review for hidden aspect and sentiment detection. It also leverages sentimental overall ratings, which often comes with online reviews, as supervision data, and can infer the semantic aspects and aspect-level sentiments that are not only meaningful but also predictive of overall sentiments of reviews. Moreover, we also develop efficient inference method for parameter estimation of SJASM based on collapsed Gibbs sampling. We evaluate SJASM extensively on real-world review data, and experimental results demonstrate that the proposed model outperforms seven well-established baseline methods for sentiment analysis tasks.

Index Terms—Sentiment analysis, aspect-based sentiment analysis, probabilistic topic model, supervised joint topic model.

1 Introduction

Online user-generated reviews are of great practical use, because: 1) They have become an inevitable part of decision making process of consumers on product purchases, hotel bookings, etc. 2) They collectively form a low-cost and efficient feedback channel, which helps businesses to keep track of their reputations and to improve the quality of their products and services. As a matter of fact, online reviews are constantly growing in quantity, while varying largely in content quality. To support users in digesting the huge amount of raw review data, many sentiment analysis techniques have been developed for past years [1].

Generally, sentiments and opinions can be analyzed at different levels of granularity. We call the sentiment expressed in a whole piece of text, e.g., review document or sentence, *overall sentiment*. The task of analyzing overall sentiments of texts is typically formulated as classification problem, e.g., classifying a review document into positive or negative sentiment. Then, a variety of machine learning methods trained using different types of indicators (features) have been employed for overall sentiment analysis [2], [3], [4], [5], [6], [7]. However, analyzing the overall sentiment expressed in a whole piece of text alone (e.g., review document), does not discover what specifically people like or dislike in the text. In reality, the fine-grained sentiments may very well tip the balance in purchase decisions. For example, savvy consumers nowadays are no longer satisfied with just overall sentiment/rating given to a product in a review; They are often eager to see why it receives that rating, which positive or negative attributes (aspects) contribute to the particular rating of the product.

Recently, there has been a growing interest in analyzing *aspect-level sentiment*, where an *aspect* means a unique semantic facet of an entity commented on in text documents, and is typically represented as a high-level hidden cluster of semantically related keywords (e.g., aspect terms). Aspect-based sentiment analysis generally consists of two major tasks, one is to detect hidden semantic *aspect* from given texts, the other is to identify fine-grained sentiments expressed towards the aspects. Probabilistic topic models, which are typically built on a basic latent Dirichlet allocation (LDA) model [8], have been used for aspect-based sentiment analysis [9], [10], [11], [12], [13], [14], [15], where the semantic *aspect* can be naturally formulated as one type of latent topics (latent variables).

To our knowledge, most majority of existing probabilistic joint topic-sentiment (or sentiment-topic) models are unsupervised or weakly/partially supervised, meaning that they primarily model user-generated text content, and have not considered overall ratings or labels of the text documents in their frameworks. As a result, though they may capture the hidden thematic structure of text data, the models cannot directly predict the overall sentiments or ratings of text documents, instead, they only rely on *document-specific sentiment distribution* to approximate the overall sentiments of documents.

Moreover, previous studies usually treat overall sentiment analysis and aspect-based sentiment analysis in isolation, and then introduce a variety of methods to analyze either overall sentiments or aspect-level sentiments, but not both. We observe that there exists naturally inter-dependency between the aspect-based and overall sentiment analysis problems. Specifically, inferring predictive hidden aspects and sentiments from text reviews can be helpful for predicting overall ratings/sentiments of reviews, while overall ratings/sentiments of text reviews can provide guidance and constraint for inferring fine-grained sentiments on the aspects from the reviews. We believe a carefully designed supervised unification model can benefit from the inter-dependency between the two problems, and support them to improve each other. It is thus important to analyze aspect-level sentiments and overall sentiments in

- *E-mail: haiz0001@ntu.edu.sg.*

one go under a unified framework.

In this work, we focus on modeling online user-generated review and overall rating pairs, and aim to identify semantic aspects and aspect-level sentiments from review texts as well as to predict overall sentiments of reviews. Generally, online reviews often come with overall ratings, for example, in the form of *one-to-five* star ratings, which can be naturally regarded as sentiment labels of the text reviews. This evidence provides us with pretty good opportunity to develop supervised joint topic model for aspect-based and overall sentiment analysis problems. In particular, instead of using *bag-of-words* representation, which is typically adopted for processing usual text data (e.g., articles), we first represent each text review as a bag of opinion pairs, where each opinion pair consists of an aspect term and corresponding opinion word in the review. We extend the basic LDA model, and construct a probabilistic joint aspect and sentiment framework to model the textual *bag-of-opinion-pairs* data. Then, on top of the probabilistic topic modeling framework, we introduce a new supervised learning layer via normal linear model to jointly capture overall rating information. In addition, we also leverage weak supervision data based on pre-compiled sentiment lexicon, which provides sentimental prior knowledge for the model. In this way, we develop a novel supervised joint aspect and sentiment model (SJASM) which is able to cope with aspect-based sentiment analysis and overall sentiment analysis in a unified framework.

Several key advantages of SJASM help it stand out in the probabilistic joint topic models to sentiment analysis: 1) SJASM can simultaneously model aspect terms and corresponding opinion words of each text review for semantic aspect and sentiment detection; 2) It exploits sentimental overall ratings as supervision data, and can infer the semantic aspects and fine-grained aspect-level sentiments that are not only meaningful but also predictive of overall sentiments of reviews; and 3) It leverages sentiment prior information, and can explicitly build the correspondence between detected sentiments (latent variables) and real-world sentiment orientations (e.g., positive or negative).

Moreover, based on the collapsed Gibbs sampling method [16], [17], we present a new efficient inference algorithm to estimate the parameters for SJASM. We use publicly available real-world review data to evaluate SJASM for three typical sentiment analysis tasks, i.e., semantic aspect detection, aspect-level sentiment identification, and overall rating/sentiment prediction. The experimental results demonstrate the superiority of SJASM over seven well-established baseline methods.

Next, this work has made the following main contributions:

- This work presents a new supervised joint topic model called SJASM, which forms the prediction for overall ratings/sentiments of reviews via normal linear model based on the inferred hidden aspects and sentiments in the reviews.
- It formulates overall sentiment analysis and aspect-based sentiment analysis in a unified framework, which allows SJASM to leverage the inter-dependency between the two problems and to support the problems to improve each other.
- It presents a detailed inference method for SJASM based on collapsed Gibbs sampling.
- This work evaluates SJASM against seven representative baseline methods, and experimentally demonstrates the benefits of SJASM over them for sentiment analysis problems.

The rest of this article is organized as follows. We present related work to sentiment analysis in Section 2, and problem definition in Section 3. Wee describe the proposed supervised joint topic model SJASM in Section 4, and derive the detailed inference procedure for the model in Section 5. Section 6 presents the experimental results of the proposed model for sentiment analysis tasks. Section 7 provides discussions on the proposed model. In Section 8, we conclude this article, and present related future directions to this work.

2 Related Work

In this section, we present related work to overall sentiment analysis and aspect-based sentiment analysis, notably the family of probabilistic topic models for the latter.

2.1 Overall Sentiment Analysis

Sentiments and opinions can be analyzed not only at different levels of granularity, but also for different types of data, e.g., user-generated review data and social media data.

2.1.1 User-generated Review Data

By formulating overall sentiment analysis as a classification problem, Pang et al. [2] built supervised models on standard *n-gram* text features to classify review documents into positive or negative sentiments. Moreover, to prevent a sentiment classifier from considering non-subjective sentences, Pang and Lee [18] used a subjectivity detector to filter out non-subjective sentences of each review, and then applied the classifier to resulting subjectivity extracts for sentiment prediction. A similar two-stage method was also proposed in [3] for document-level sentiment analysis. A variety of features (indicators) have been evaluated for overall sentiment classification tasks. Zhao et al. [4] employed a conditional random fields based model to incorporate contextual dependency and label redundancy constraint features for sentence-level sentiment classification, while Yang and Cardie [7] incorporated lexical and discourse constraints at intra-/inter-sentence level via a similar model for the problem. Liu and Seneff [19] exploited linguistic adverbial and negation features via a parse-and-paraphrase method to predict the sentiments of product reviews. Paltoglou and Thelwall [20] studied information retrieval related features and weighting schemes for sentiment classification. Different types of embeddings learned from review data have been used for sentiment analysis. Maas et al. [6] first proposed an unsupervised probabilistic model to learn word embeddings, and then, based on the embeddings of words appearing in given reviews, they trained a supervised classification model to deal with the sentiment analysis tasks at both document and sentence levels. Socher et al. [21] exploited hierarchical structures and compositional semantics via a recursive auto-encoder model to create sentence embeddings. Then, they built a supervised classification model on the sentence embeddings for sentiment prediction. Besides textual review data, Tang et al. [22] leveraged continuous user and product embeddings learned via unified user-product neural network model for sentiment classification of review documents.

2.1.2 Social Media Data

Sentiment analysis of social media data, such as tweets, blogs, and forums, has attracted extensive attention, which can be perhaps viewed as sentiment analysis at document or sentence level.

Abbasi et al. [23] first selected stylistic and syntactic features via entropy weighted genetic method, and then, they trained a supervised classification model on the features for sentiment prediction in Web forums. To analyze overall sentiments of blog (and review) documents, Melville et al. [5] incorporated background/prior lexical knowledge based on a pre-compiled sentiment lexicon into a supervised pooling multinomial text classification model. Hu et al. [24] combined sentimental consistency and emotional contagion with supervised learning for sentiment classification in micro-blogging. As a matter of fact, different from user-generated review data, which often come with labeled overall ratings (e.g., one-to-five star ratings), social media domain has been suffering from the scarcity of high-quality labeled data. Paltoglou and Thelwall [25] proposed an unsupervised lexicon-based approach for sentiment classification on Twitter, MySpace, and Digg. Tan et al. [26] leveraged social relationship data in addition to limited labeled data, and developed a semi-supervised method to predict the sentiments expressed in text tweets. Liu et al. [27] extracted two sets of text and non-text features on Twitter networks, and used a two-view co-training method for semi-supervised learning to classify sentiments of tweet data.

In addition, sentiments and opinions can be also analyzed at word or phrase level, where the objective is to predict the sentiment polarities of opinion words or phrases [28], [29], [30]. However, sentiment analysis at document, sentence, or word level alone does not discover what exactly people like or dislike in the texts. Nowadays, people are no longer satisfied with just overall sentiments expressed in a whole piece of text, and moreover, they may care about what specific *aspects* of the opinionated entity are mentioned, and which particular sentiment orientations (e.g., positive or negative) have been expressed towards the aspects in the text.

2.2 Aspect-based Sentiment Analysis

Recently, there has been a growing interest in aspect-based sentiment analysis. It has been previously known as *feature-specific* sentiment analysis, where the *feature* is different from the *aspect*, and generally corresponds to a particular *aspect term* that is explicitly commented on in a text document.

2.2.1 Structural Tagging Methods

By formulating feature-specific sentiment analysis as a structural labeling problem, Jin et al. [31] developed a lexicalized hidden Markov models based method to integrate linguistic factors (e.g., POS-tags) and contextual clues of words into the sequential learning process for recognizing features (aspect terms), opinion words, and opinion orientations from reviews. Similarly, Li et al. [32] relied on a sequential tagging model based on conditional random fields (CRFs) to deal with the fine-grained review analysis and summarization. Jakob and Gurevych [33] also used the CRFs model for single-domain and cross-domain feature extraction problem. One limitation of the aforementioned models is that they need large-scale fine-grained labeled/tagged review data for model building, which are very difficult to come by in reality.

2.2.2 Linguistic Methods

Unsupervised linguistic methods rely on developing syntactic rules or dependency patterns to cope with fine-grained sentiment analysis problem. Qiu et al. [34] proposed a syntactic parsing based double propagation method for feature-specific sentiment analysis. Based on dependency grammar [35], they first defined eight syntactic rules, and employed the rules to recognize pair-wise word dependency for each review sentence. Then, given opinion word seeds, they iteratively extracted more opinion words and the related features, by relying on the identified syntactic dependency relations. They inferred the sentiment polarities on the features via a heuristic contextual evidence based method during the iterative extraction process. Wu et al. [36] presented a phrase dependency parsing method to recognize features, opinion expressions, as well as the dependency relations between them. Linguistic approaches are domain-independent, in the sense that the syntactic rules or dependency patterns developed in a domain can be readily applied to a different domain. However, the approaches tend to suffer from: 1) the limited coverage of the manually defined syntactic rules, and 2) the colloquial nature of real-life reviews, which typically contain informal content or grammatically incorrect sentences.

2.2.3 Corpus Statistics Methods

Corpus statistics methods rely on mining frequent statistical patterns to address sentiment analysis problems. The methods are somewhat resistant to informal language of online text documents, provided that the given text corpus is suitably large. Hu and Liu [37] proposed an association rule mining [38] approach (ARM) to discover the frequently mentioned nouns or noun phrases in product reviews as potential features. They then utilized compactness and redundancy pruning methods to filtered out the irrelevant product features. Furthermore, Popescu and Etzioni [39] employed a point-wise mutual information model [40] to prune the frequent but invalid feature candidates, leading to improved feature extraction performance compared to ARM. Hai et al. [41] proposed to employ corpus statistics association to measure pairwise word dependency, and introduced a generalized association based bootstrapping method for extracting features and their associated opinion words from reviews.

However, all the aforementioned methods do not group extracted synonymous or semantically related keywords (e.g., features) into concise high-level semantic *aspect* clusters or *aspects*. There is perhaps redundancy in the sentiment and opinion summarization results, as it is common that different people often use a variety of words to express the same aspect. For example, all the specific features, "screen", "LCD", and "display", which are explicitly mentioned in reviews, refer to the same aspect "screen" in cellphone review domain. A separate step of categorization or clustering [42], [43] may be applied, but it will result in additional accumulation of errors.

2.3 Probabilistic Topic Models

Probabilistic topic models, which are typically built on basic latent Dirichlet allocation (LDA) model [8], have been widely used for aspect-based sentiment analysis.

Titov and McDonald [44] introduced a multi-aspect sentiment model to analyze aspect-level sentiments from user-generated reviews. The model assumption, i.e., individual aspect-related ratings are present in reviews, may lead to the limited use in reality, since a large number of online reviews are not annotated with the semantic aspects and aspect-specific opinion ratings by online users. Lin and He [45] extended the LDA model by designing a sentiment layer, and introduced a joint sentiment-topic model (JST) for sentiment analysis. Then, Lin et al. [11] extended the JST model by incorporating sentiment prior knowledge based on pre-compiled sentiment lexicons, and introduced a weakly supervised joint sentiment-topic model. One limitation of their model is that it cannot directly predict overall sentiments of review documents. Wang et al. [46] developed a two-stage approach for latent aspect rating analysis (LARA). They first identified semantic aspects via bootstrapping algorithm. They then inferred the fine-grained sentiment ratings on the aspects via a partially supervised latent rating regression model. Similarly, their model cannot predict overall sentiments/ratings of review documents. This is because they treated overall rating information as constraint to infer the weights of hidden aspects. Jo and Oh [9] proposed a weakly supervised aspect and sentiment unification model (ASUM) to detect sentiments towards different aspects in a unified framework. But the model assumption, i.e., each review sentence contains exactly one aspect/topic, is often violated, when the model is applied to real-life complicated review documents. Moghaddam and Ester [10] introduced an unsupervised aspect-sentiment LDA model to identify latent aspects and their sentiment labels from online product reviews. One shortcoming of the unsupervised model is that the correspondence between detected hidden sentiments (latent variables) and real-world sentiment labels is not specified. Kim et al. [12] proposed a hierarchical aspect-sentiment model to discover hidden hierarchical structure of aspect-level sentiments from unlabeled online reviews. In addition to user-generated text reviews, Dermouche et al. [13] leveraged time stamp data, and developed an unsupervised time-aware topic-sentiment graphical model for analyzing topic-sentiment evolution over time, while Yang et al. [14] exploited the demographic information of reviewers (user meta-data), and proposed a partially supervised user-aware sentiment-topic model for aspect-based sentiment analysis problem. Rahman and Wang [15] built an unsupervised hidden topic-sentiment model to capture topic coherence and sentiment consistency in text reviews for recognizing latent aspects and corresponding sentiments.

As far as we know, most majority of existing probabilistic joint topic-sentiment (sentiment-topic) models are unsupervised or weakly/partially supervised, meaning that they often model user-generated text content, and do not consider sentimental overall rating or label of text documents in their frameworks. Moreover, the overall sentiment analysis and aspect-based sentiment analysis problems are typically handled in isolation in previous studies. In contrast, in this work, we focus on modeling user-generated review and overall rating pair data, and propose a new supervised joint topic model named SJASM to deal with the two sentiment analysis problems in one go under a unified framework. One key advantage of SJASM over previous sentiment analysis techniques is that it can leverage the inter-dependency between the two problems, and support the problems to boost each other. In addition, online user-generated reviews often come with overall ratings, i.e., sentiment labels, it is thus effortless to construct labeled review data for learning the proposed model.

As an extended work, it is related to but different from the original work [47], as shown below. 1) This work deals with a different problem, i.e., *sentiment analysis*, while the original work focuses on *review quality evaluation* problem. 2) This work introduces a generalization of the normal linear model used for modeling overall rating response. 3) The SJASM model presented in this work leverages sentimental overall ratings of reviews and lexical prior information as supervision data. 4) A detailed collapsed Gibbs sampling procedure is derived for parameter estimation of SJASM.

3 Problem Definition

Since we analyze user-generated review data, we first provide the definitions of the terminologies commonly used in the sentiment analysis of user reviews.

Aspect Term: An *aspect term* t, also known as *feature* [37] or *explicit feature* [48], indicates a specific attribute or component word of an opinionated entity (product), which typically appears as noun or noun phrase in review text. For instance, the noun "voice" is an aspect term in the audio CD review, "She has a powerful *voice* that is different from most others."

Opinion Word: An *opinion word* o, also called *sentiment word* [49], refers to the word used to express subjectivity or sentiments, and typically appears as adjective in review documents. For example, the word "powerful" is recognized as an opinion word from the aforementioned example review.

Opinion Pair: An *opinion pair* $op = \langle t, o \rangle$ is simply defined as a pair of aspect term t and corresponding opinion word o extracted from a given review document. For instance, one opinion pair $op = \langle voice, powerful \rangle$ can be recognized from the example review above. The extracted opinion pairs would constitute the input to our sentiment analysis system.

Aspect: An *aspect* a, or semantic aspect, refers to a unique facet that corresponds to a ratable attribute or component of an opinionated entity. In our setting, it is formulated as latent variable, and is typically represented as a hidden cluster of semantically related aspect terms or opinion words. For instance, one can detect a ratable semantic aspect *appearance* from the following audio CD reviews, "The appearance and design are so beautiful." and "It is really good-looking."

Sentiment: A *sentiment* s, or *opinion*, refers to the semantic orientation and degree (strength) of satisfaction on a reviewed entity or its aspect in a review text. Positive semantic orientation indicates praise (e.g., "good"), while negative semantic orientation indicates criticism (e.g., "bad") [28]. In our setting, *sentiment* is formulated as a latent variable, and refers to a hidden semantic cluster of opinion words which share the same sentimental polarity.

Overall Rating: The *overall rating* r indicates the degree of sentiment demonstrated in a whole review document.

Then, given an entity (e.g., product) from a category, there is a collection of M review documents on the entity, $D = \{d_1, d_2, \ldots, d_M\}$. Each review d_m can be reduced to a list of N opinion pairs: $d_m = \{\langle t_1, o_1\rangle, \langle t_2, o_2\rangle, \ldots, \langle t_N, o_N\rangle\}$, where each opinion pair consists of an aspect term t_n and corresponding opinion word o_n in the review. We aim to deal with three sentiment analysis tasks as follows.

- Semantic aspect detection. This task aims at detecting hidden semantic aspects of an opinionated entity from the given review documents, where each aspect would be represented in the form of a hidden semantic cluster.
- Aspect-level sentiment identification. For this task, the aim is to identify fine-grained semantic sentiment orientations, e.g., *positive* or *negative*, expressed towards each detected semantic aspect.
- Overall rating/sentiment prediction. Given an unlabeled review, we will form the prediction for the overall sentimental rating by employing a carefully designed regression procedure over the inferred hidden aspects and aspect-level sentiments via the fitted model.

4 METHODOLOGY

4.1 Overview

We model online user-generated review and overall rating pairs, and aim to identify semantic aspects and aspect-level sentiments from review texts as well as to predict overall sentiments of reviews.

User-generated reviews are different from ordinary text documents. For example, when people read a product review, they often care about which specific aspects of the product are commented on, and what sentiment orientations (e.g., positive or negative) have been expressed on the aspects. Instead of employing bag-of-words representation, which is typically adopted for processing usual text documents, we represent each review in an intuitive form of opinion pairs, where each opinion pair consists of an aspect term and related opinion word in the review. Probabilistic topic models, notably latent Dirichlet allocation (LDA) [8], have been widely used for analyzing semantic topical structure of text data. Based on the basic LDA, we introduce an additional aspect-level sentiment identification layer, and construct a probabilistic joint aspect and sentiment framework to model the textual bag-of-opinion-pair data. Online user-generated reviews often come with overall ratings (sentiment labels), which provides us with great flexibility to develop supervised unification topic model. Then, on top of the constructed probabilistic framework, we introduce a new supervised learning layer via normal linear model to jointly model the overall rating data. Thus, we propose a novel supervised joint aspect and sentiment model (SJASM), which can cope with the overall and aspect-based sentiment analysis problems in one go under a unified framework.

4.2 Supervised Joint Aspect and Sentiment Model

We make the following assumptions about our proposed SJASM model:

- The generation for aspect-specific sentiments depends on the aspects. This means that we first generate latent aspects, on which we subsequently generate corresponding sentiment orientations.
- The generation for aspect terms depends on the aspects, while the generation for opinion words relies on the sentiment orientations and semantic aspects. The formulation is intuitive. For example, to generate an opinion word "beautiful", we need to know not only its sentiment orientation, e.g., *positive*, but also the related semantic aspect, e.g., *appearance*.
- The generation for overall ratings of reviews depends on the aspect-level sentiments in the reviews.

Based on the assumptions, the SJASM model generates a review document and its overall rating in the following way. It first draws hidden semantic aspects conditioned on document-specific aspect distribution; Then, it draws the sentiment orientations on the aspects conditioned on the per document aspect-specific sentiment distribution; Next, it draws each opinion pair, which contains an aspect term and corresponding opinion word, conditioned on both aspect and sentiment specific word distributions; Lastly, it draws the overall rating response based on the generated aspects and sentiments in the review document.

The graphical representation of the proposed SJASM model is shown in Figure 1. The notations used in the SJASM model are listed in Table 1. Under SJASM, a review

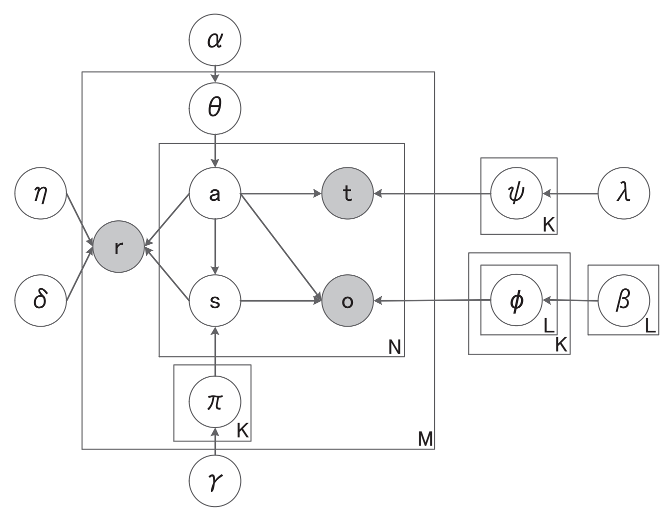

Fig. 1. Graphical representation of SJASM. The boxes refers to plates that indicate replicates. The outer plate refers to review documents, while the inner plate refers to the repeated selection of latent aspects and sentiment orientations as well as aspect terms and opinion words within each review document.

TABLE 1
Notations used in SJASM

M	Number of review documents in a corpus
N	Number of opinion pairs in a review
K	Number of semantic aspects
L	Number of semantic sentiments
t_{mn}	Aspect term of nth opinion pair in review d_m
o_{mn}	Opinion word of nth opinion pair in d_m
a_{mn}	Aspect assignment to term t_{mn} and word o_{mn}
s_{mn}	Sentiment assignment to opinion word o_{mn}
r_m	Overall rating response of review d_m
θ	Dirichlet distribution over aspects
π	Dirichlet distribution over sentiments
ψ	Dirichlet distribution over aspect words
ϕ	Dirichlet distribution over opinion words
α	Hyperparameter for aspect distribution θ
γ	Hyperparameter for sentiment distribution π
λ	Hyperparameter for aspect word distribution ψ
β	Hyperparameter for opinion word distribution ϕ
η	Overall rating response parameter
δ	Overall rating response parameter
U	Vocabulary of unique aspect words
V	Vocabulary of unique opinion words
$\boldsymbol{a}^{\neg i}$	All aspect assignments except for a_i
$\boldsymbol{s}^{\neg i}$	All sentiment assignments except for s_i
$N_{m,k}$	Count of words in d_m assigned to aspect k
$N_{m,k,l}$	Count of words in d_m assigned to k and l
$N_{k,u}$	Count of aspect word u assigned to aspect k
N_k	Total count of aspect words assigned to aspect k
$N_{k,l,v}$	Count of opinion word v assigned to k and l
$N_{k,l}$	Total count of opinion words assigned to k and l

document d_m and its overall rating r_m are generated from the following formal process:

- For each aspect $k \in \{1, \ldots, K\}$
  1) Draw aspect word distribution $\psi_k \sim$ Dir(λ).
  2) For each sentiment orientation $l \in \{1, \ldots, L\}$
     a) Draw opinion word distribution $\phi_{kl} \sim$ Dir(β_l).
- For each review d_m and its overall rating r_m
  1) Draw aspect distribution $\theta_m \sim$ Dir(α).
  2) For each aspect k under review r_m
     a) Draw sentiment distribution $\pi_{mk} \sim$ Dir(γ).
  3) For an opinion pair $\langle t_{mn}, o_{mn}\rangle, n \in \{1, \ldots, N\}$
     a) Draw aspect assignment $a_{mn} \sim$ Mult(θ_m).
     b) Draw sentiment assignment $s_{mn} \sim$ Mult($\pi_{ma_{mn}}$).
     c) Draw aspect term $t_{mn} \sim$ Mult($\psi_{a_{mn}}$).
     d) Draw opinion word $o_{mn} \sim$ Mult($\phi_{a_{mn}s_{mn}}$).
  4) Draw overall rating response $r_m \sim$ N($\eta^T \bar{z}_m$, δ).

Note that $\bar{z}_m$ refers to the empirical frequencies of hidden variables (latent aspects and sentiments) in the review document d_m, and is defined as

$$\bar{z}_m = \frac{1}{C}\sum_{n=1}^{N}(a_{mn} \times (\omega^T \times s_{mn})),$$

where ω consists of normalization coefficients on latent sentiment variables, and C means normalization constant. Under the framework of SJASM, overall rating response r_m of review d_m is drawn from a normal linear model N($\eta^T \bar{z}_m$, δ), where η and δ refer to rating response parameters. In this normal linear model, the covariates corresponds to the empirical frequencies of hidden aspects and sentiments $\bar{z}_m$, and η represents the regression coefficients on the empirical frequencies.

Moreover, the aforementioned normal linear model can be viewed as a special case of a generalized linear model (GLM) [50]. In GLM, a response variable r is assumed to be generated from a particular distribution in the exponential dispersion family, parameterized by ρ and δ,

$$p(r|\rho, \delta) = h(r, \delta)exp\{\frac{\rho r - A(\rho)}{\delta}\}, \tag{1}$$

where ρ is usually called *natural parameter* and is related to the mean of the distribution, and δ, known as *dispersion parameter*, is related to the variance of the distribution. Following [51], if we set $\rho = \eta^T \bar{z}_m$ for a given review document d_m, and substitute the term into Equation 1, then we can obtain Equation 2 of GLM for generating overall rating r_m of the document, i.e., $r_m \sim$ GLM($\bar{z}_m$, η, δ),

$$p(r_m|\bar{z}_m, \eta, \delta) = h(r_m, \delta)exp\{\frac{(\eta^T \bar{z}_m)r_m - A(\eta^T \bar{z}_m)}{\delta}\}. \tag{2}$$

The generalization via GLM can provide flexibility in modeling different types of overall rating response of a review document, whose distribution can be shown in the exponential dispersion form. In reality, a large range of probability distributions are applicable under this framework, including the normal distribution for real-valued response and binomial distribution for binary response, etc. Each of the distributions corresponds to a specific selection of $h(r_m, \delta)$ and $A(\eta^T \bar{z}_m)$. In particular, it is straightforward to verify for normal distribution that

$$h(r_m, \delta) = \frac{1}{\sqrt{2\pi\delta}}exp\{\frac{-{r_m}^2}{2\delta}\}$$

and

$$A(\eta^T \bar{z}_m) = \frac{(\eta^T \bar{z}_m)^2}{2}.$$

Obviously, the usual parameters of normal distribution μ and σ^2 correspond to $\eta^T \bar{z}_m$ and δ, respectively.

The novel formulation behind the proposed SJASM model actually agrees well with intuitions. Generally, different products have diverse lists of aspects, e.g., attributes or components. The utility quality of individual product aspects could be different, and may result in different evaluations and opinions on the aspects. Overall experiences and sentiments on the products would be formed or regressed on the product aspects and their associated evaluations expressed in the reviews. The regression coefficients reflect the relative contributions of the fine-grained aspect-specific sentiments. Furthermore, given user-generated review and rating pair data, the labeled overall ratings of review documents can be leveraged as supervision knowledge. They thus provide useful guidance and constraint on the procedure of inferring the meaningful and predictive hidden aspects and sentiments.

In addition to overall rating data, SJASM also leverages a pre-compiled sentiment lexicon as weak supervision information, which not only benefits semantic sentiment analysis, but also provides explicit correspondence between latent sentiment variables and real-world sentiment orientations (e.g., positive or negative).

5 Inference and Estimation

In this section, we derive detailed inference procedure for parameter estimation of the proposed SJASM model.

5.1 Model Inference

The target of inference is to evaluate the posterior distribution of the hidden variables given each review document and its rating:

$$\begin{aligned} &p(\boldsymbol{a}, \boldsymbol{s} \mid \boldsymbol{t}, \boldsymbol{o}, \boldsymbol{r}, \alpha, \gamma, \lambda, \beta, \eta, \delta) \\ &= \frac{p(\boldsymbol{a}, \boldsymbol{s}, \boldsymbol{t}, \boldsymbol{o}, \boldsymbol{r} \mid \alpha, \gamma, \lambda, \beta, \eta, \delta)}{p(\boldsymbol{t}, \boldsymbol{o}, \boldsymbol{r} \mid \alpha, \gamma, \lambda, \beta, \eta, \delta)}. \end{aligned} \quad (3)$$

Since exact inference for the posterior distribution is intractable, based on the collapsed Gibbs sampling algorithm [16], [17], we present a new efficient inference method for SJASM.

In particular, for each opinion pair $\langle t_{mn}, o_{mn} \rangle$ with the index $i = (m, n)$, we are interested in deriving the following full conditional distribution:

$$p(a_i = k, s_i = l \mid \boldsymbol{a}^{\neg i}, \boldsymbol{s}^{\neg i}, \boldsymbol{t}, \boldsymbol{o}, \boldsymbol{r}, \alpha, \gamma, \lambda, \beta, \eta, \delta),$$

where the quantities $\boldsymbol{a}^{\neg i}$ and $\boldsymbol{s}^{\neg i}$ indicate the assignments of aspects and sentiment orientations to all aspect terms and opinion words in the corpus except for the assignments a_i and s_i for the aspect term and opinion word at the position i, respectively.

We expand the conditional distribution as follows:

$$\begin{aligned} &p(a_i = k, s_i = l \mid \boldsymbol{a}^{\neg i}, \boldsymbol{s}^{\neg i}, \boldsymbol{t}, \boldsymbol{o}, \boldsymbol{r}, \alpha, \gamma, \lambda, \beta, \eta, \delta) \\ &= \frac{p(\boldsymbol{a}, \boldsymbol{s}, \boldsymbol{t}, \boldsymbol{o}, \boldsymbol{r} \mid \alpha, \gamma, \lambda, \beta, \eta, \delta)}{p(\boldsymbol{a}^{\neg i}, \boldsymbol{s}^{\neg i}, \boldsymbol{t}, \boldsymbol{o}, \boldsymbol{r} \mid \alpha, \gamma, \lambda, \beta, \eta, \delta)} \\ &\propto p(\boldsymbol{a}, \boldsymbol{s}, \boldsymbol{t}, \boldsymbol{o}, \boldsymbol{r} \mid \alpha, \gamma, \lambda, \beta, \eta, \delta). \end{aligned} \quad (4)$$

Under SJASM, the joint distribution of aspect terms, opinion words, the assignments of aspects and sentiments, and overall ratings can be factored as follows:

$$\begin{aligned} &p(\boldsymbol{a}, \boldsymbol{s}, \boldsymbol{t}, \boldsymbol{o}, \boldsymbol{r} \mid \alpha, \gamma, \lambda, \beta, \eta, \delta) \\ &= p(\boldsymbol{a} \mid \alpha) \cdot p(\boldsymbol{s} \mid \boldsymbol{a}, \gamma) \cdot p(\boldsymbol{t} \mid \boldsymbol{a}, \lambda) \\ &\cdot p(\boldsymbol{o} \mid \boldsymbol{a}, \boldsymbol{s}, \beta) \cdot p(\boldsymbol{r} \mid \boldsymbol{a}, \boldsymbol{s}, \eta, \delta). \end{aligned} \quad (5)$$

For the first term in Equation 5, by integrating out θ we yield:

$$p(\boldsymbol{a} \mid \alpha) = \prod_m \frac{\Gamma(\sum_k \alpha_k)}{\prod_k \Gamma(\alpha_k)} \frac{\prod_k \Gamma(N_{m,k} + \alpha_k)}{\Gamma(N + \sum_k \alpha_k)}. \quad (6)$$

where $N_{m,k}$ indicates the number of times the words in the review document d_m are assigned to aspect k, N is the total number of words in the document, and $\Gamma(x)$ refers to the Gamma function.

Integrating out π for the second term in Equation 5, we obtain:

$$p(\boldsymbol{s} \mid \boldsymbol{a}, \gamma) = \prod_m \prod_k \frac{\Gamma(L\gamma)}{\Gamma(\gamma)^L} \frac{\prod_l \Gamma(N_{m,k,l} + \gamma)}{\Gamma(N_{m,k} + L\gamma))}. \quad (7)$$

where $N_{m,k,l}$ is the number of times the words in document d_m are assigned to the aspect k and sentiment orientation l.

We integrate out ψ for the third term and obtain:

$$p(\boldsymbol{t} \mid \boldsymbol{a}, \lambda) = \prod_k \frac{\Gamma(|U|\lambda)}{\Gamma(\lambda)^{|U|}} \frac{\prod_u \Gamma(N_{k,u} + \lambda)}{\Gamma(N_k + |U|\lambda)}. \quad (8)$$

where $N_{k,u}$ is the number of times the unique aspect word u in vocabulary U is assigned to the aspect k across all documents, and N_k indicate the sum of $N_{k,u}$ across all the unique aspect words in vocabulary.

Integrating out ϕ for the fourth term, we yield:

$$\begin{aligned} &p(\boldsymbol{o} \mid \boldsymbol{a}, \boldsymbol{s}, \beta) \\ &= \prod_k \prod_l \frac{\Gamma(\sum_v \beta_{l,v})}{\prod_v \Gamma(\beta_{l,v})} \frac{\prod_v \Gamma(N_{k,l,v} + \beta_{l,v})}{\Gamma(N_{k,l} + \sum_v \beta_{l,v})}. \end{aligned} \quad (9)$$

where $N_{k,l,v}$ is the count of unique opinion word v in vocabulary V assigned to sentiment l and aspect k, and $N_{k,l}$ is the sum of $N_{k,l,v}$ across all the unique opinion words in vocabulary.

We expand the last term in Equation 5 as follows:

$$\begin{aligned} &p(\boldsymbol{r} \mid \boldsymbol{a}, \boldsymbol{s}, \eta, \delta) \\ &= \prod_m \frac{1}{\sqrt{2\pi\delta}} exp(-\frac{(r_m - \eta^T \bar{z}_m)^2}{2\delta}). \end{aligned} \quad (10)$$

Then, we substitute all the equations 6, 7, 8, 9, and 10 into Equation 5. We cancel out the terms that are independent of the aspect term t_{mn} and opinion word o_{mn}, as well as the aspect and sentiment assignments a_{mn} and s_{mn}. We yield the full conditional distribution for the opinion pair $\langle t_{mn}, o_{mn} \rangle$ with index $i = (m, n)$ as follows:

$$\begin{aligned} &p(a_i = k, s_i = l \mid \boldsymbol{a}^{\neg i}, \boldsymbol{s}^{\neg i}, \boldsymbol{t}, \boldsymbol{o}, \boldsymbol{r}, \alpha, \gamma, \lambda, \beta, \eta, \delta) \\ &\propto \frac{N_{m,k}^{\neg i} + \alpha_k}{N^{\neg i} + \sum_{k'} \alpha_{k'}} \cdot \frac{N_{m,k,l}^{\neg i} + \gamma}{N_{m,k}^{\neg i} + L\gamma} \cdot \frac{N_{k,u}^{\neg i} + \lambda}{N_k^{\neg i} + |U|\lambda} \\ &\cdot \frac{N_{k,l,v}^{\neg i} + \beta_{l,v}}{N_{k,l}^{\neg i} + \sum_{v'} \beta_{l,v'}} \cdot \frac{1}{\sqrt{2\pi\delta}} exp(-\frac{(r_m - \eta^T \bar{z}_m^{\neg i})^2}{2\delta}). \end{aligned} \quad (11)$$

Note the subscript $-i$ in a quantity indicates the exclusion of the corresponding data at the position i. For instance, $N_{k,u}^{\neg i}$ represents the number of times the unique aspect word u in vocabulary is assigned to aspect k across all documents, excluding the particular aspect term and corresponding aspect assignment at the index i.

5.2 Parameter Estimation

Next, we compute the model parameters of SJASM based on the samples obtained via the collapsed Gibbs sampling method.

We compute per document aspect distribution as follows:

$$\theta_{m,k} = \frac{N_{m,k} + \alpha_k}{N + \sum_{k'=1}^{K} \alpha_{k'}}. \quad (12)$$

Then, we compute per document aspect specific sentiment distribution:

$$\pi_{m,k,l} = \frac{N_{m,k,l} + \gamma}{N_{m,k} + L\gamma}. \quad (13)$$

The aspect word distribution is computed as follows:

$$\psi_{k,u} = \frac{N_{k,u} + \lambda}{N_k + |U|\lambda}. \quad (14)$$

The opinion word distribution is computed:

$$\phi_{klv} = \frac{N_{k,l,v} + \beta_{l,v}}{N_{k,l} + \sum_{v'=1}^{|V|} \beta_{l,v'}}. \quad (15)$$

In addition, we choose asymmetric Dirichlet prior α for the document aspect distribution θ, and estimate the hyperparameter α by employing a fixed-point iteration scheme [52]. We exploit sentimental prior information by using asymmetric hyperparameter β for the opinion word distribution ϕ. The prior knowledge comes from a publicly available lexicon named MPQA Subjectivity Lexicon[1]. In our experimental setting, we consider two sentiment orientations ($L = 2$), positive and negative. The elements of the hyperparameter β that correspond to positive opinion words are specified as high values for positive orientation, i.e., β_{lv} =0.95, and low values for negative orientation, i.e., β_{lv} =0.05, respectively. The values of the hyperparameter β corresponding to negative opinion words are specified inversely. Following previous work [11], [17], we choose symmetric priors γ and λ for the document aspect specific sentiment distribution π and aspect word distribution ψ, and set them as $1/L$ and 0.01, respectively.

We follow Blei and McAuliffe [51] and approximately estimate the overall rating response parameters η and δ. Let Z be a $M \times L$ matrix, whose rows correspond to the vectors $\bar{z}_m^T$. Then, the regression coefficients η can be approximated as follows:

$$\hat{\eta} \approx (Z^T Z)^{-1} Z^T r, \tag{16}$$

where the vector r consists of the overall ratings for all the reviews. The parameter δ can be estimated:

$$\hat{\delta} \approx \frac{1}{M}\{r^T r - r^T Z (Z^T Z)^{-1} Z^T r\}. \tag{17}$$

6 EXPERIMENTS

6.1 Data Sets

We evaluated our proposed SJASM model using publicly available user-generated review data from three categories, i.e., game (video game), CD (audio CD), and hotel. The game and CD reviews[2] were collected from Amazon[3], and the hotel reviews[4] were collected from TripAdvisor[5]. Table 2 shows some statistics of the review data sets.

TABLE 2
Statistics of Review Data Sets

Category	Game	CD	Hotel
# of reviews in corpus	2,599	1,632	1,367
# of words in corpus	554,496	292,060	696,327
Vocabulary size	23,809	13,886	21,785
Average # of words/review	213	178	509

6.2 Review Data Preprocessing

In preprocessing the review data, we parsed all the review documents in each data set using the well-known *Stanford Parser*[6], which has been previously shown to perform well in working out grammatical dependency structures of texts in practice. Following [47], we recognized the opinion pairs $op = \langle t, o \rangle$ from the parsed review documents by relying on the following dependency relations defined in the parser, *nsubj* (i.e., nominal subject), *dobj* (i.e., direct object), and *amod* (i.e., adjectival modifier). For example, from this CD review, "He's got a unique sound.", we can recognize an adjectival phrase "amod(sound, unique)" via the adjectival modifier dependency relation *amod*, and then straightforwardly extract $\langle sound, unique \rangle$ as an opinion pair. Next, we expanded the extracted list of opinion pairs by applying two additional dependency relations *neg* (i.e., negation modifier) and *conj* (i.e., conjunct). For example, in this CD review, "The exterior and design are very beautiful.", we can first identify an opinion pair $\langle exterior, beautiful \rangle$ via employing the dependency relation *nsubj*, then, we can recognize a conjunction phrase "conj(exterior, design)" via the conjunct relation *conj*, and thus identify another opinion pair $\langle design, beautiful \rangle$ by employing the related dependency relations.

We note that, for each of the processed game, CD, and hotel review sets, the extracted opinion pairs of review documents and the corresponding overall ratings constitute our labeled training data. Then, we held out 20% of the review data for testing purpose, and trained the proposed model on the remaining 80% of the data.

6.3 Baselines and Parameter Settings

We evaluated the proposed SJASM model for three typical sentiment analysis tasks: 1) Semantic aspect detection, 2) Aspect-level sentiment identification, and 3) Overall rating/sentiment prediction. We compared SJASM with seven well-established representative benchmark models, including supervised latent Dirichlet allocation model (sLDA) [51], weakly supervised joint sentiment-topic model (JST) [11], aspect-sentiment unification model (ASUM) [9], latent aspect rating analysis model (LARA) [46], supervised pooling multinomial model (Pooling) [5], SVM based sentiment classifier (SVM) [2], and unsupervised lexicon based method (Lexicon) [53].

The models SJASM, sLDA, JST, and ASUM were evaluated for all the three tasks. The LARA model was evaluated for the semantic aspect detection and aspect-level sentiment identification, because LARA was developed for latent aspect rating analysis and cannot be used for predicting overall sentiment/rating of a review. The last three baselines Pooling, SVM, and Lexicon were evaluated only for overall rating/sentiment prediction, as they cannot discover the hidden thematic structure of review data.

In our experiments, we followed previous studies and determined the values of the parameters for respective baseline methods as follows. In particular, following [51] for the sLDA model, we fixed α to $1/K$ (K: number of topics), and initialized $\beta_{1:K}$ to randomly perturbed uniform topics, σ^2 to the sample variance of the response, and η to a grid on $[-1, 1]$ in increments of $2/K$. For the JST model, we followed [11] and set the symmetric prior $\beta = 0.01$, the symmetric prior $\gamma = (0.05 * L)/S$ (L: average document length; S: number of sentiment labels), and the asymmetric prior α was learned from data using maximum-likelihood estimation. Following [9], we fixed, for ASUM, the symmetric hyperparameters $\alpha = 0.1$ and $\gamma = 1.0$, and set the

1. http://mpqa.cs.pitt.edu/lexicons/subj_lexicon
2. http://liu.cs.uic.edu/download/data
3. http://www.amazon.com
4. http://sifaka.cs.uiuc.edu/wang296/Data/index.html
5. http://www.tripadvisor.com
6. http://nlp.stanford.edu/software/lex-parser.html

elements of the hyperparameter β to be 0.0 for positive (negative) seed words given negative (positive) sentiment labels and 0.001 for all other words. With following [46], the parameters μ, Σ, σ^2, and β for the LARA model were estimated from the review data using maximum likelihood estimator. Following [5] for the Pooling method, we used the sigmoid weighting scheme to compute the weights α, and set the polarity level $r = 100$. We followed [2] and used the well-known SVM^{light} package[7], where all related parameters were set to the default values, to build the SVM based sentiment classifier.

6.4 Gold Standard Data and Evaluation Metrics

User-generated reviews from *Amazon* or *TripAdvisor* come with overall ratings in the form of *one-to-five* ratings. It is thus straightforward to obtain the gold standard data for the third evaluation task, i.e., overall rating/sentiment prediction. In order to conduct fair and reliable comparisons with existing baseline models, we followed [9], [11], and transformed the gold standard overall ratings into binary sentiment labels. Specifically, the overall ratings that are larger than the neutral score, for example, 3.0 for *one-to-five* rating scale, were converted to *positive* labels, while the ratings that are lower than (or equal to) the neutral score were converted to *negative* labels. In the same way, we transformed the predicted overall ratings of SJASM and sLDA into the binary sentiment labels. We used *accuracy* (A) to evaluate the performance of overall sentiment prediction for SJASM and baselines. The accuracy measures the proportion of the sentiment labels that are predicted correctly in the held-out test set, defined as follows:

$$A = \frac{1}{n}\sum_{i=1}^{n} c_i,$$

where n is the size of the test set, and c_i is *1* if the label of review document is predicted correctly, and 0 otherwise.

Since the fine-grained aspect and aspect-level sentiment information is not available in our collected data sets, we have to manually generate the gold standard data. To create the gold standard for hidden aspect detection, we first extracted a list of aspect keywords that appear frequently in each review data set, then, we asked two human judges to simply partition the list into K word clusters, one for each aspect. When there is a conflict between them, an additional judge helps to make a final decision.

For our experiment, we specified the number of hidden aspects as $K = 5$ for easy comparison, and used *Rand index* (R) [54], a standard measure of the similarity between two data clusterings/partitions, to measure the performance for semantic aspect detection. In particular, let S be a set of n elements $S = \{e_1, \ldots, e_n\}$, P be a partition of S into p subsets $P = \{P_1, \ldots, P_p\}$, and Q be another partition of S into q subsets $Q = \{Q_1, \ldots, Q_q\}$, The Rand index R is then computed as follows:

$$R = \frac{n_1 + n_2}{n_1 + n_2 + n_3 + n_4},$$

where n_1 is the number of pairs of elements in S that are in the same set in P and in the same set in Q, n_2 is the number of pairs of elements in S that are in different sets in P and in different sets in Q, n_3 is the number of pairs of elements in S that are in the same set in P but in different sets in Q, and n_4 is the number of pairs of elements in S that are in different sets in P but in the same set in Q.

Next, to create the gold standard for aspect-level sentiment identification, the human judges then manually recognized the sentiments on the detected aspects in each test review document. We followed previous work [9], [11], and considered two sentiment orientations ($L = 2$), i.e., *positive* and *negative*. We used the *accuracy* metric to measure the performance of SJASM and baselines for this task.

6.5 Semantic Aspect Detection

First, we evaluated the proposed SJASM model for hidden semantic aspect detection against baseline models sLDA, JST, ASUM, and LARA. As all the evaluated models produce soft clustering results, we determined the aspect clusters that have the highest probability as the identified aspects for given aspect keywords. The baselines Pooling, SVM, and Lexicon were excluded in this evaluation, because they cannot discover the underlying topical structure of text review data.

TABLE 3
Rand Index of Different Models for Semantic Aspect Detection (*K=5*)

	SJASM	sLDA	JST	ASUM	LARA
Game	83.21%	75.44%	73.65%	70.35%	66.03%
CD	80.19%	73.21%	72.09%	68.91%	67.50%
Hotel	78.07%	72.11%	70.37%	69.29%	63.68%
Average	80.49%	73.59%	72.04%	69.52%	65.74%

Table 3 shows the Rand Index of the evaluated models on different data sets, the higher, the better. SJASM achieves the best Rand index score of 83.21% on the game review data set, which is 7.77%, 9.56%, 12.86%, and 17.18% higher than that of baseline models sLDA, JST, ASUM, and LARA, respectively. It also attains the highest Rand index value of 80.19% on the CD category, which is 6.98%, 8.1%, 11.28%, and 12.69% better than sLDA, JST, ASUM, and LARA. On hotel reviews, SJASM again results in the best Rand index of 78.07%, which is 5.96%, 7.7%, 8.78%, and 14.39% better than sLDA, JST, ASUM, and LARA, respectively. For each evaluated model, we also showed the average aspect detection performance across all the data sets.

The proposed SJASM model results in improved aspect identification results compared to baseline models for the following reasons: 1) SJASM leverages overall ratings as supervision data to guide the process of identifying meaningful semantic aspects; and 2) It represents each review document in the form of opinion pairs, and simultaneously exploits individual pairs of aspect terms and corresponding opinion words for hidden aspect detection. In contrast, the baseline methods sLDA, JST, ASUM, and LARA cannot gain from such special design.

Table 4 lists example aspects detected via the proposed SJASM model on the game, CD, and hotel review data sets. We showed the factual keywords, i.e., aspect terms, clustered in the hidden aspects “gameplay”, “sound”, and “staff” in the columns one, three, and five, while we listed

7. http://svmlight.joachims.org

the sentimental aspect keywords, i.e., opinion words, in the column two, four, and six, respectively. The top five opinion words are positive, and the subsequent five opinion words are negative. SJASM benefits from modeling reviews in the form of bag-of-opinion-pairs, and thus, does not mix up the aspect terms and associated opinion words in each aspect cluster.

TABLE 4
Aspects Detected via SJASM on Different Review Data Sets (*K=5*)

"gameplay" @ Game		"sound" @ CD		"staff" @ Hotel	
game	great	voice	powerful	staff	friendly
gameplay	good	vocals	beautiful	people	helpful
graphics	easy	sound	strong	service	nice
story	fun	songs	emotional	smile	attentive
storyline	enjoyable	talent	haunting	waiters	polite
characters	boring	piano	eerie	management	terrible
music	stupid	range	weird	bartenders	rude
controls	bad	quality	raw	waitress	poor
line	annoying	ballads	spooky	server	flip
plot	hard	melody	distorted	employees	fresh

From Table 4, we can see that, for each example aspect in the given review domain, the factual aspect terms are coherent and specific, while the clustered opinion words are semantically related to the aspect. For example, the semantic aspect "gameplay" for the game domain groups the factual aspect terms, such as "story", "storyline", and "controls", and the opinion words, such as "easy", "enjoyable", and "boring". The factual and sentimental keywords reflect in part the way that the game is designed and the skills that players may need in order to play it, e.g., "controls" and "hard". They are clearly indicative of the "gameplay" aspect. Then aspect "sound" for CD category clusters the factual aspect terms like "voice", "vocals", and "sound", and the semantically related opinion words like "powerful", "beautiful", and "distorted". The aspect "staff" also groups the set of semantically related keywords in the hotel review domain.

6.6 Aspect-level Sentiment Identification

Next, we evaluated SJASM and related baseline models for aspect-specific sentiment identification, i.e., identifying fine-grained sentiments towards semantic aspects. We relied on the inferred per document aspect-specific sentiment distribution (π) for this task. Note that we simply used the predicted overall sentiments of reviews for sLDA, since there is no aspect-level sentiment analysis module/layer in sLDA.

TABLE 5
Accuracy for Aspect-level Sentiment Identification on Different Review Data Sets (*K=5*, *L=2*)

	SJASM	sLDA	JST	ASUM	LARA
Game	76.24%	59.32%	68.25%	64.54%	63.64%
CD	70.95%	54.41%	63.98%	61.00%	62.96%
Hotel	69.44%	52.38%	63.16%	58.82%	58.15%
Average	72.21%	55.37%	65.13%	61.45%	61.58%

Table 5 shows the accuracy of the evaluated models for aspect-level sentiment identification. SJASM results in the best accuracy of 76.24% on the game data set, which is 16.92%, 7.99%, 11.7%, and 12.6% higher than that of sLDA, JST, ASUM, and LARA, respectively. On the CD review data set, it again achieves the highest accuracy of 70.95%, which is 16.54%, 6.97%, 9.95%, and 7.99% better than sLDA, JST, ASUM, and LARA. SJASM also achieves the best accuracy of 69.44% on the hotel review domain, which is 17.06%, 6.28%, 10.62%, and 11.29% better than sLDA, JST, ASUM, and LARA, respectively.

SJASM outperforms all the baselines for aspect-specific sentiment identification on the three review domains. SJASM introduces the normal linear model to jointly leverage the overall ratings of review documents as supervision data in the unification framework, and provides guidance and constraint on identifying fine-grained sentiment orientations towards the aspects in the reviews. SJASM also leverages weak prior knowledge that comes from pre-compiled sentiment lexicon.

In contrast, although the two baseline models JST and ASUM also leverage the same weak supervision data based on sentiment lexicon, they do not consider the overall rating data evidence in their frameworks, and only relying on text review content to infer the fine-grained sentiments. They thus lose the competition with the proposed model. Though including the overall ratings of reviews in its modeling structure, LARA exploits the overall rating data to estimate the weights put on different semantic aspects, and loses out to the SJASM model for aspect-level sentiment analysis. The baseline model sLDA leverages the overall ratings, but it does not have the aspect-specific sentiment identification layer in the modeling structure. Simply adopting the predicted overall sentiment of a review as the fine-grained sentiments on the aspects mentioned in the review leads to the worst results. It is very common in reality, due to different utility quality of aspects of an entity, users comment positively on some aspects in a review, meanwhile show negative sentiments on other aspects in the same review.

6.7 Overall Sentiment Prediction

We evaluated the performance of predicting overall sentiments of reviews for SJASM and six representative baseline methods, including supervised methods sLDA, Pooling, and SVM, weakly supervised methods JST and ASUM, and one unsupervised method Lexicon. LARA was excluded from the evaluation, as it cannot predict overall sentiments/ratings of text reviews.

The proposed SJASM model formulates overall rating (sentiment) prediction task as a specialized regression problem. Given an unlabeled review document, SJASM infers its hidden topical and sentimental structure using a fitted model, and then forms the prediction by regressing the overall rating response on the empirical frequencies of underlying aspects and sentiments in the review. The reported results were averaged over five runs for each review domain. We showed the average accuracy with error bars versus the number of latent aspects. Note that only one accuracy value was shown for each of the baselines Pooling, SVM, and Lexicon, since they cannot model hidden thematic structure of text review data.

Figure 2 plots overall sentiment prediction results on the game data set. SJASM outperforms all the benchmark models sLDA, Pooling, SVM, JST, ASUM, and Lexicon across

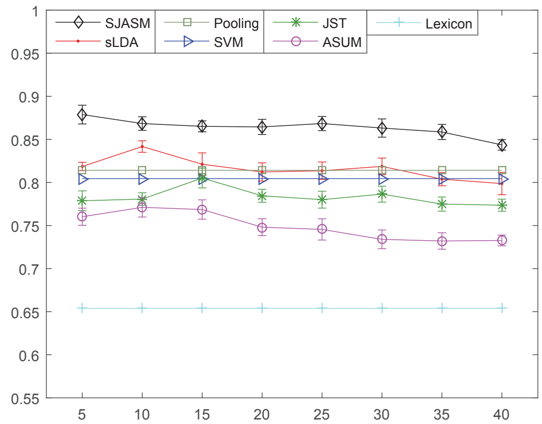

Fig. 2. Overall sentiment prediction accuracy versus aspect number on the game data set.

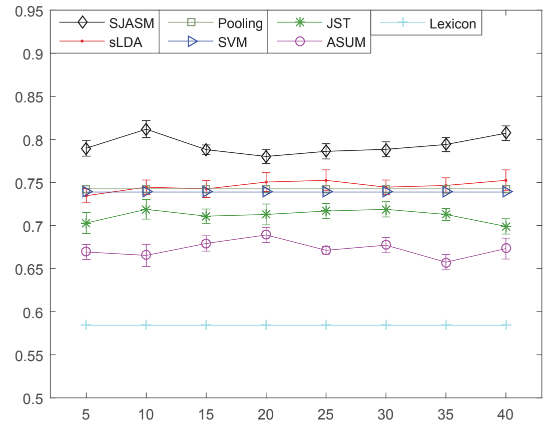

Fig. 4. Overall sentiment prediction accuracy versus aspect number on the hotel data set.

all eight aspect numbers (from 5 to 40). SJASM achieves the best overall sentiment prediction accuracy of 87.88% (at aspect number $K = 5$), which is 3.71%, 6.47%, 7.43%, 7.37%, 10.77%, , and 22.50% higher than that of sLDA (at aspect number 10), Pooling, SVM, JST (15), and ASUM (10), and Lexicon, respectively.

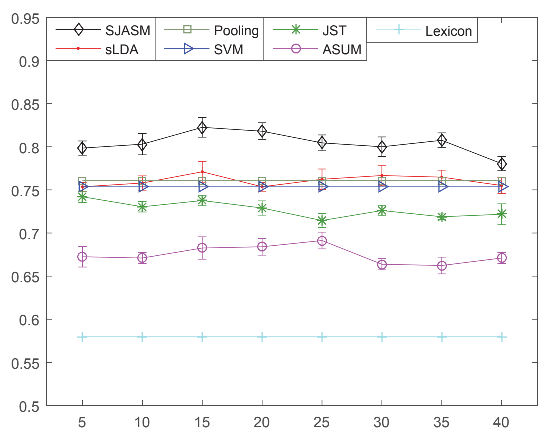

Fig. 3. Overall sentiment prediction accuracy versus aspect number on the CD data set.

Figure 3 plots the overall sentiment prediction performance on the CD data set. Again, SJASM outperforms all the baseline models. As the number of aspects grows to $K = 15$, SJASM achieves the best accuracy of 82.26%, which is 5.16%, 6.17%, 6.90%, 8.06%, 13.13%, and 24.29% better than that of sLDA (15), Pooling, SVM, JST (5), ASUM (25), and Lexicon, respectively. Figure 4 plots the overall sentiment prediction results on the hotel data set. The best accuracy of SJASM is 81.19% (at aspect number 10), which is 5.94%, 6.93%, 7.3%, 9.31%, 12.28%, and 22.77% higher than that of sLDA (25), Pooling, SVM, JST (10), ASUM (20), and Lexicon, respectively.

The unsupervised lexicon-based method (Lexicon) results in the worst overall sentiment prediction performance compared to all other methods. Because the Lexicon method tends to suffer from informal expressions, such as colloquial language and misspelled words, which are frequently mentioned in online user-generated reviews, but may not be included in the sentiment lexicon. The two weakly supervised models JST and ASUM gain from well designed probabilistic joint modeling structure, which helps improve the sentiment prediction performance compared to the Lexicon method. However, they lose the competition with all the fully supervised models for predicting overall sentiments of review documents. One interpretation would be that both JST and ASUM only model the text review content, and do not consider the overall rating labels of reviews in their frameworks, thus, they cannot directly predict the overall sentiments of review documents, instead, they can only approximate the sentiments by using inferred document-specific sentiment distributions.

The baseline Pooling relies on incorporating sentimental background knowledge into its supervised learning framework, and performs a little better than the SVM sentiment classifier, which was built using standard text unigram features. But both of them perform worse compared to the supervised topic modeling methods. One illustration may be that supervised topic models benefit from supervised dimensionality reduction, while both Pooling and SVM do not model meaningful topical structure of review data, and thus cannot gain from this.

Though sLDA performs better than other baselines for overall sentiment prediction, it loses out to the proposed SJASM model. The superiority of SJASM over sLDA can be attributed to new specialized design for sentiment analysis: 1) SJASM copes with aspect-level and overall sentiment analysis tasks in one go under a unified framework, which allows it to leverage the inter-dependency between the two tasks and to help them boost each other; 2) It jointly detect the semantic aspects and sentiments under supervision of overall rating information, which enables it to infer which positive or negative aspects are much predictive for overall sentiment prediction; and 3) In addition to supervised overall rating information, SJASM leverages prior sentimental knowledge, which comes from a pre-compiled sentiment lexicon, sLDA cannot benefit from this.

7 Discussions

We develop supervised joint aspect and sentiment model (SJASM) to analyze overall and aspect-level sentiments for online user-generated review data, which often come with labeled overall rating information. Note that this work does not aim to deal with the problem of sentiment analysis on social media data, e.g., tweets or blogs, where the overall ratings or sentimental labels usually are not available. Then, if we try to apply the proposed model SJASM to social media data for sentiment analysis, one choice could be to manually annotate the overall ratings or sentimental labels for social media text data.

User-generated review data are different from usual textual articles. When people read reviews, they typically concern themselves with what aspects of an opinionated entity are mentioned in the reviews, and which sentiment orientations are expressed towards the aspects. Thus, instead of using traditional bag-of-words representation, we reduce each text review as a bag of opinion pairs, where each opinion pair contains an aspect term and related opinion word appearing in the review. Specifically, we parsed all the text reviews in each data set using the well-known *Stanford Parser*, and then straightforwardly relied on the syntactic dependency patterns to recognize the opinion pairs from the review texts. As a separate preprocessing step, several other methods, which were specially developed for extracting aspect terms and corresponding opinion words from reviews [34], [55], can be perhaps used for generating the bag-of-opinion-pairs representation. It's true that better opinion pair extraction results would be beneficial for the proposed model SJASM to achieve improved performance for sentiment analysis tasks.

The proposed SJASM model belongs to the family of generative probabilistic topic modeling approaches to sentiment analysis. SJASM is able to model the hidden thematic structure of text review data. Thus, similar to other unsupervised or weakly supervised joint topic-sentiment (sentiment-topic) models, it can rely on per document-specific sentiment distribution to approximate the overall ratings or sentiments of text reviews. However, according to the experimental results, the performance is not as good as that achieved by leveraging new supervised normal linear model. Under the supervised unified framework of SJASM, we can infer hidden semantic aspects and sentiments that are predictive of overall ratings of text reviews. Then, to form the prediction for overall sentiments of reviews, we directly regress the sentiment response on the inferred latent aspects and sentiments in the reviews. It is the specialized design of SJASM that makes big difference.

8 Conclusions

In this work, we focus on modeling online user-generated review data, and aim to identify hidden semantic aspects and sentiments on the aspects, as well as to predict overall ratings/sentiments of reviews. We have developed a novel supervised joint aspect and sentiment model (SJASM) to deal with the problems in one go under a unified framework. SJASM treats review documents in the form of opinion pairs, and can simultaneously model aspect terms and their corresponding opinion words of the reviews for semantic aspect and sentiment detection. Moreover, SJASM also leverages overall ratings of reviews as supervision and constraint data, and can jointly infer hidden aspects and sentiments that are not only meaningful but also predictive of overall sentiments of review documents. We conducted experiments using publicly available real-world review data, and extensively compared SJASM with seven well-established representative baseline methods. For semantic aspect detection and aspect-level sentiment identification problems, SJASM outperforms all the generative benchmark models, sLDA, JST, ASUM, and LARA. As for overall sentiment prediction, SJASM again outperforms the six benchmark methods sLDA, Pooling, SVM, JST, ASUM, and Lexicon.

Online user-generated reviews are often associated with location or time-stamp information. For future work, we will extend the proposed model by modeling the meta-data to cope with the spatio-temporal sentiment analysis of online reviews. Probabilistic topic modeling approaches to sentiment analysis often requires the number of latent topics to be specified in advance of analyzing review data. Another interesting future direction of our work is to develop Bayesian nonparametric model, which can automatically estimate the number of latent topics from review data, and can also allow the number of the topics to increase as new review examples appear.

Acknowledgments

This work was supported in part by a grant awarded by a Singapore MOE AcRF Tier 2 Grant (ARC30/12), a Singapore MOE AcRF Tier 1 Grant (RG 66/12), and a National Research Foundation, Prime Ministers Office, Singapore under its IDM Futures Funding Initiative. This work was done while Z. Hai was working at Nanyang Technological University, Singapore.

References

[1] B. Liu, "Sentiment analysis and opinion mining," *Synthesis Lectures on Human Language Technologies*, vol. 5, no. 1, pp. 1–167, May 2012.

[2] B. Pang, L. Lee, and S. Vaithyanathan, "Thumbs up?: sentiment classification using machine learning techniques," in *Proceedings of the ACL-02 conference on Empirical methods in natural language processing - Volume 10*, ser. EMNLP'02. Stroudsburg, PA, USA: Association for Computational Linguistics, 2002, pp. 79–86.

[3] V. Ng, S. Dasgupta, and S. M. N. Arifin, "Examining the role of linguistic knowledge sources in the automatic identification and classification of reviews," in *Proceedings of the COLING/ACL on Main Conference Poster Sessions*, ser. COLING-ACL '06. Stroudsburg, PA, USA: Association for Computational Linguistics, 2006, pp. 611–618.

[4] J. Zhao, K. Liu, and G. Wang, "Adding redundant features for crfs-based sentence sentiment classification," in *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, ser. EMNLP '08. Stroudsburg, PA, USA: Association for Computational Linguistics, 2008, pp. 117–126.

[5] P. Melville, W. Gryc, and R. D. Lawrence, "Sentiment analysis of blogs by combining lexical knowledge with text classification," in *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ser. KDD'09. New York, NY, USA: ACM, 2009, pp. 1275–1284.

[6] A. L. Maas, R. E. Daly, P. T. Pham, D. Huang, A. Y. Ng, and C. Potts, "Learning word vectors for sentiment analysis," in *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies - Volume 1*, ser. HLT'11. Stroudsburg, PA, USA: Association for Computational Linguistics, 2011, pp. 142–150.

[7] B. Yang and C. Cardie, "Context-aware learning for sentence-level sentiment analysis with posterior regularization," in *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics, ACL 2014, June 22-27, 2014, Baltimore, MD, USA, Volume 1: Long Papers*, 2014, pp. 325–335.

[8] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent dirichlet allocation," *J. Mach. Learn. Res.*, vol. 3, pp. 993–1022, March 2003.

[9] Y. Jo and A. H. Oh, "Aspect and sentiment unification model for online review analysis," in *Proceedings of the fourth ACM international conference on Web search and data mining*, ser. WSDM'11. New York, NY, USA: ACM, 2011, pp. 815–824.

[10] S. Moghaddam and M. Ester, "Ilda: Interdependent lda model for learning latent aspects and their ratings from online product reviews," in *Proceedings of the 34th International ACM SIGIR Conference on Research and Development in Information Retrieval*, ser. SIGIR'11, 2011, pp. 665–674.

[11] C. Lin, Y. He, R. Everson, and S. Ruger, "Weakly supervised joint sentiment-topic detection from text," *IEEE Transactions on Knowledge and Data Engineering*, vol. 24, no. 6, pp. 1134–1145, Jun. 2012.

[12] S. Kim, J. Zhang, Z. Chen, A. Oh, and S. Liu, "A hierarchical aspect-sentiment model for online reviews," in *Proceedings of the Twenty-Seventh AAAI Conference on Artificial Intelligence*, ser. AAAI'13. AAAI Press, 2013, pp. 526–533.

[13] M. Dermouche, J. Velcin, L. Khouas, and S. Loudcher, "A joint model for topic-sentiment evolution over time," in *2014 IEEE International Conference on Data Mining*, Dec 2014, pp. 773–778.

[14] Z. Yang, A. Kotov, A. Mohan, and S. Lu, "Parametric and non-parametric user-aware sentiment topic models," in *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval*, ser. SIGIR '15. New York, NY, USA: ACM, 2015, pp. 413–422.

[15] M. M. Rahman and H. Wang, "Hidden topic sentiment model," in *Proceedings of the 25th International Conference on World Wide Web*, ser. WWW '16. Republic and Canton of Geneva, Switzerland: International World Wide Web Conferences Steering Committee, 2016, pp. 155–165.

[16] S. Geman and D. Geman, "Stochastic relaxation, gibbs distributions, and the bayesian restoration of images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 6, no. 6, pp. 721–741, 1984.

[17] T. L. Griffiths and M. Steyvers, "Finding scientific topics," in *Proceedings of the National Academy of Science*, vol. 101, Jan 2004, pp. 5228–5235.

[18] B. Pang and L. Lee, "A sentimental education: sentiment analysis using subjectivity summarization based on minimum cuts," in *Proceedings of the 42nd Annual Meeting on Association for Computational Linguistics*, ser. ACL'04. Stroudsburg, PA, USA: Association for Computational Linguistics, 2004.

[19] J. Liu and S. Seneff, "Review sentiment scoring via a parse-and-paraphrase paradigm," in *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 1 - Volume 1*, ser. EMNLP '09. Stroudsburg, PA, USA: Association for Computational Linguistics, 2009, pp. 161–169.

[20] G. Paltoglou and M. Thelwall, "A study of information retrieval weighting schemes for sentiment analysis," in *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, ser. ACL '10. Stroudsburg, PA, USA: Association for Computational Linguistics, 2010, pp. 1386–1395.

[21] R. Socher, J. Pennington, E. H. Huang, A. Y. Ng, and C. D. Manning, "Semi-supervised recursive autoencoders for predicting sentiment distributions," in *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, ser. EMNLP '11. Stroudsburg, PA, USA: Association for Computational Linguistics, 2011, pp. 151–161.

[22] D. Tang, B. Qin, and T. Liu, "Learning semantic representations of users and products for document level sentiment classification," in *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*. Beijing, China: Association for Computational Linguistics, July 2015, pp. 1014–1023.

[23] A. Abbasi, H. Chen, and A. Salem, "Sentiment analysis in multiple languages: Feature selection for opinion classification in web forums," *ACM Trans. Inf. Syst.*, vol. 26, no. 3, pp. 12:1–12:34, Jun. 2008.

[24] X. Hu, L. Tang, J. Tang, and H. Liu, "Exploiting social relations for sentiment analysis in microblogging," in *Proceedings of the Sixth ACM International Conference on Web Search and Data Mining*, ser. WSDM '13. New York, NY, USA: ACM, 2013, pp. 537–546.

[25] G. Paltoglou and M. Thelwall, "Twitter, myspace, digg: Unsupervised sentiment analysis in social media," *ACM Trans. Intell. Syst. Technol.*, vol. 3, no. 4, pp. 66:1–66:19, Sep. 2012.

[26] C. Tan, L. Lee, J. Tang, L. Jiang, M. Zhou, and P. Li, "User-level sentiment analysis incorporating social networks," in *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ser. KDD '11. New York, NY, USA: ACM, 2011, pp. 1397–1405.

[27] S. Liu, X. Cheng, F. Li, and F. Li, "Tasc:topic-adaptive sentiment classification on dynamic tweets," *IEEE Transactions on Knowledge and Data Engineering*, vol. 27, no. 6, pp. 1696–1709, June 2015.

[28] P. D. Turney and M. L. Littman, "Measuring praise and criticism: Inference of semantic orientation from association," *ACM Trans. Inf. Syst.*, vol. 21, no. 4, pp. 315–346, Oct. 2003.

[29] T. Wilson, J. Wiebe, and P. Hoffmann, "Recognizing contextual polarity in phrase-level sentiment analysis," in *Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing*, ser. HLT'05. Stroudsburg, PA, USA: Association for Computational Linguistics, 2005, pp. 347–354.

[30] A. Agarwal, F. Biadsy, and K. R. Mckeown, "Contextual phrase-level polarity analysis using lexical affect scoring and syntactic n-grams," in *Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics*, ser. EACL'09. Stroudsburg, PA, USA: Association for Computational Linguistics, 2009, pp. 24–32.

[31] W. Jin, H. H. Ho, and R. K. Srihari, "Opinionminer: A novel machine learning system for web opinion mining and extraction," in *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ser. KDD'09. New York, NY, USA: ACM, 2009, pp. 1195–1204.

[32] F. Li, C. Han, M. Huang, X. Zhu, Y.-J. Xia, S. Zhang, and H. Yu, "Structure-aware review mining and summarization," in *Proceedings of the 23rd International Conference on Computational Linguistics*, ser. COLING'10. Stroudsburg, PA, USA: Association for Computational Linguistics, 2010, pp. 653–661.

[33] N. Jakob and I. Gurevych, "Extracting opinion targets in a single- and cross-domain setting with conditional random fields," in *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*, ser. EMNLP'10. Stroudsburg, PA, USA: Association for Computational Linguistics, 2010, pp. 1035–1045.

[34] G. Qiu, B. Liu, J. Bu, and C. Chen, "Opinion word expansion and target extraction through double propagation," *Comput. Linguist.*, vol. 37, pp. 9–27, 2011.

[35] L. Tesniere, *Elements de Syntaxe Structurale*. Paris: Librairie C. Klincksieck, 1959.

[36] Y. Wu, Q. Zhang, X. Huang, and L. Wu, "Phrase dependency parsing for opinion mining," in *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 3 - Volume 3*, ser. EMNLP'09. Stroudsburg, PA, USA: Association for Computational Linguistics, 2009, pp. 1533–1541.

[37] M. Hu and B. Liu, "Mining and summarizing customer reviews," in *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, ser. KDD'04. New York, NY, USA: ACM, 2004, pp. 168–177.

[38] R. Agrawal, T. Imielinski, and A. Swami, "Mining association rules between sets of items in large databases," in *Proceedings of the 1993 ACM SIGMOD international conference on Management of data*, ser. SIGMOD'93. New York, NY, USA: ACM, 1993, pp. 207–216.

[39] A.-M. Popescu and O. Etzioni, "Extracting product features and opinions from reviews," in *Proceedings of the Conference on Human Language Technology and Empirical Methods in Natural Language Processing*, ser. HLT'05. Stroudsburg, PA, USA: Association for Computational Linguistics, 2005, pp. 339–346.

[40] K. W. Church and P. Hanks, "Word association norms, mutual information and lexicography," in *Proceedings of the 27th Annual Conference of the Association of Computational Linguistics*. New Brunswick, NJ: Association for Computational Linguistics, 1989, pp. 76–83.

[41] Z. Hai, K. Chang, and G. Cong, "One seed to find them all: mining opinion features via association," in *Proceedings of the 21st ACM international conference on Information and knowledge management*, ser. CIKM'12. New York, NY, USA: ACM, 2012, pp. 255–264.

[42] H. Guo, H. Zhu, Z. Guo, X. Zhang, and Z. Su, "Product feature categorization with multilevel latent semantic association," in *Proceeding of the 18th ACM conference on Information and knowledge management*, ser. CIKM'09. New York, NY, USA: ACM, 2009, pp. 1087–1096.
[43] Z. Zhai, B. Liu, H. Xu, and P. Jia, "Clustering product features for opinion mining," in *Proceedings of the fourth ACM international conference on Web search and data mining*, ser. WSDM'11. New York, NY, USA: ACM, 2011, pp. 347–354.
[44] I. Titov and R. McDonald, "A joint model of text and aspect ratings for sentiment summarization," in *Proceedings of ACL-08: HLT*. Columbus, Ohio: Association for Computational Linguistics, June 2008, pp. 308–316.
[45] C. Lin and Y. He, "Joint sentiment/topic model for sentiment analysis," in *Proceedings of the 18th ACM conference on Information and knowledge management*, ser. CIKM'09. New York, NY, USA: ACM, 2009, pp. 375–384.
[46] H. Wang, Y. Lu, and C. Zhai, "Latent aspect rating analysis on review text data: A rating regression approach," in *Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ser. KDD'10. New York, NY, USA: ACM, 2010, pp. 783–792.
[47] Z. Hai, G. Cong, K. Chang, W. Liu, and P. Cheng, "Coarse-to-fine review selection via supervised joint aspect and sentiment model," in *Proceedings of the 37th International ACM SIGIR Conference on Research and Development in Information Retrieval*, ser. SIGIR '14. New York, NY, USA: ACM, 2014, pp. 617–626.
[48] Z. Hai, K. Chang, J.-J. Kim, and C. C. Yang, "Identifying features in opinion mining via intrinsic and extrinsic domain relevance," *IEEE Transactions on Knowledge and Data Engineering*, vol. 99, no. PrePrints, p. 1, 2013.
[49] J. Yi, T. Nasukawa, R. Bunescu, and W. Niblack, "Sentiment analyzer: Extracting sentiments about a given topic using natural language processing techniques," in *Proceedings of the Third IEEE International Conference on Data Mining*, ser. ICDM'03. Washington, DC, USA: IEEE Computer Society, 2003, pp. 427–434.
[50] P. McCullagh and J. A. Nelder, *Generalized linear models*. CRC press, 1989, vol. 37.
[51] D. M. Blei and J. D. McAuliffe, "Supervised topic models," in *Proceedings of the 21st Annual Conference on Neural Information Processing Systems - Volume 7*, Vancouver, Canada, 2007, pp. 121–128.
[52] T. Minka, "Estimating a dirichlet distribution," in *Technical report, MIT*, 2000.
[53] M. Taboada, J. Brooke, M. Tofiloski, K. Voll, and M. Stede, "Lexicon-based methods for sentiment analysis," *Comput. Linguist.*, vol. 37, no. 2, pp. 267–307, Jun. 2011.
[54] W. M. Rand, "Objective criteria for the evaluation of clustering methods," *Journal of the American Statistical association*, vol. 66, no. 336, pp. 846–850, 1971.
[55] Z. Hai, K. Chang, G. Cong, and C. C. Yang, "An association-based unified framework for mining features and opinion words," *ACM Trans. Intell. Syst. Technol.*, vol. 6, no. 2, pp. 26:1–26:21, Mar. 2015.

Zhen Hai received the PhD degree in computer science and engineering from Nanyang Technological University, Singapore. He is a research scientist at the Institute for Infocomm Research, A*STAR, Singapore. His research interests include sentiment analysis and opinion mining, natural language processing, information retrieval, data mining, and machine learning.

Gao Cong received the PhD degree from the National University of Singapore. He is an associate professor with Nanyang Technological University. Before that, he worked with Aalborg University, Microsoft Research Asia, and the University of Edinburgh. His current research interests include geospatial and textual data management, mining social network and social media, data mining, and large scale data analytics.

Kuiyu Chang received the BEng degree from National Taiwan University, the MS degree from the University of Hawaii at Manoa, and the PhD degree from the University of Texas at Austin, all in electrical/computer engineering. He is currently chief scientist at LinkSure China Holding Private Limited, Singapore. Prior to that, he was a professor of computer engineering at Nanyang Technological University for 10 years, specializing in data mining, sentiment analysis, and open source technologies. From 2000 to 2002, he was a member of technical staff at Interwoven (now HP Autonomy Division), USA. He is a co-recipient of three international best paper awards (PAKDD 2012, ISI 2005, and Motorola 1996) and one merit paper award (IAENG 2012). He has consulted for IT companies in China, Singapore, and Malaysia on data-mining and open-source software adoption.

Peng Cheng received the BEng and MS degrees from Shandong University, and the PhD degree from Nanyang Technological University, all in computer science and engineering. She is a research fellow in the School of Computer Science and Engineering, Nanyang Technological University, Singapore. Her research interests include digital geometry processing, geometric modeling, discrete differential geometry, computer graphics, human computer interaction, and machine learning.

Chunyan Miao is a professor in the School of Computer Science and Engineering at Nanyang Technological University (NTU). She is the director of the NTU-UBC Joint Research Centre of Excellence in Active Living for the Elderly (LILY). Prior to joining NTU, she was an instructor and postdoctoral fellow at the School of Computing, Simon Fraser University, Canada. Her research focus is on studying the cognitive and social characteristics of intelligent agents in multi-agent and distributed AI/CI systems, such as trust, emotions, motivated learning, ecological and organizational behavior. She has made significant contributions in the integration of the above research into emerging technologies such as interactive digital media (e.g., virtual world, social networks, and massively multi-player game), cloud computing, mobile communication, and humanoid robots.