

Integrating Topic, Sentiment and Syntax for Modeling Online Review

Rui Xie, Chunping Li

School of Software
Tsinghua University
Beijing, China
{harryxse, cli}@tsinghua.edu.cn

Qiang Ding, Li Li

Shannon Lab
Huawei Technologies, LTD
Beijing, China
{q.ding, jollylili.li}@huawei.com

Abstract— The problem of analyzing online product reviews has drawn much interest of researchers. In this paper, we propose a novel probabilistic modeling framework based on *Latent Dirichlet Allocation* (LDA), which can reveal the latent aspect and sentiment of the review simultaneously. Unlike other topic models which only consider the text itself of online review, we firstly combine the *Part-of-Speech* (POS) tag into the model. We further propose three *Tag Sentiment Aspect Models* (TSA) to integrate the syntax information into modeling. The experiments show that our models are able to achieve a promising result not only on sentiment classification but on extraction of aspects of different sentiments.

Keywords—topic model; sentiment analysis; tag sentiment aspect model; online review analysis.

I. INTRODUCTION

Nowadays, the development of Web 2.0 [3] makes convenient for customers to express their experience with products. Websites like Amazon.com and Epinions.com offer a platform for people to praise or criticize the target product. As a result, large amount of product reviews can be easily got from the Internet. These reviews are useful resource to help the customer to make decisions whether or not to buy a specific product. By browsing these reviews, people learn the good or the bad aspects of specific product. On the other hand, not only the customers, but the product designers also pay more and more attentions to these reviews. However, facing the overwhelming amount of reviews of products, no one can read the reviews piece by piece. There is an urgent need for the approach to obtaining useful and hidden information from larger review corpus. From the perspective of designers, given a product and its reviews, the aspects these reviews talk about and what the customers' overall attitude towards these aspects are the most important issues.

There are two problems during this process. First is topic identification, for identifying the aspects of the product the reviews talk about. The other one is sentiment analysis, to determine the sentiment label (positive, negative) of opinion toward specific aspect. They are challenging, not only for the large volumes of reviews, but for unstructured property of the plain text.

Topic identification, also known as aspect discovery, has been studied for a long time. In the past, two ways were

usually used to extract aspect: filtering way and expanding way. The filtering way is to firstly extract a set of frequent Noun Phrases (NP) as candidate aspects, and then filter out the candidates which are less possible to be an aspect [1][2]. The expanding way is to firstly give an aspect list as basic knowledge and then expand the original list by using various expanding methods [3]. Sentiment analysis, also known as opinion mining, aims at using automated techniques to identify semantic orientation in texts. There are many works dedicated to classify the whole document or review into positive, neutral, and negative one [4][5]. Nevertheless, the sentiment of specific aspects is usually more useful than the overall rate. Other works focus on sentiment classification on the word/phrase level [6][7], but the word's sentiment polarity is dependent on topic or domain. Modeling the sentiment along with aspect/topic is required to make the result more informative.

The proposed models in this paper tackle this problem. The *Tag Sentiment Aspect Models*, extended from *LDA* [8], can model the aspects and sentiment of online reviews simultaneously. To our best knowledge, not much work can do this except [9][10][11][12]. Our models have several differences and improvements compared to existing works: (1) *TSA* models are to incorporate the syntax information into the hierarchical Bayesian model. (2) *TSA* models are fully unsupervised while some existing works need labeled data to train. (3) *TSA* models are domain independent. By integrating different domain prior information, *TSA* models can be applied to different domains.

The way *TSA* models integrating syntax information is based on the assumption that different words in sentences play different roles. A word can appear in a sentence for several reasons. It can play a role of syntactic function, and it can play a role of semantic content [13].

The rest of the paper is organized as follows. In Section II, we discuss the related works. Section III describes our proposed three *TSA* models and corresponding inference. In Section IV we show the experimental setup and give the evaluation of the model and discussion of the results in Section V. We have the conclusion and the future work in Section VI.

II. RELATED WORK

There are two major directions to discover the hidden aspect and sentiment in reviews. One direction is to apply

traditional natural language processing techniques to do text mining for reviews. For aspect discovery, the Noun Phrase (NP) detection is a widely used technique. Hu and Liu [2] used POS tagging to find noun phrase and selected frequent nouns as aspect candidates. A filtering method is applied to these candidates to generate real aspects. In [1], the similar approach is used to discover hidden aspects but besides NP, text fragments in the sentence level are as well used for generating aspects. Besides POS tagging, linguistic rules are also used to identify product feature/aspect. Turney [6] manually designed several linguistic rules to identify feature, like ‘JJ + NN + (feature)’ and ‘RB + VB + (feature)’, etc.

Unsupervised methods often need a lexicon to decide the word’s orientation. The lexicon is built by expanding from a seed list. Using lexicon, the scheme for scoring the overall sentiment of sentence or review is well designed. Lu and Zhai [14] proposed a context-aware method for constructing the lexicon to adapt for different domains. They utilized general-purpose sentiment lexicon, thesaurus, the corpus’ sentiment rating information and linguistic heuristics to reassign sentiment score to the vocabulary. The reassigned vocabulary comprises the new domain dependent lexicon. There existed some works on building general-purpose sentiment lexicon as well. The famous lexicon is *SentiWordnet* [20], which is an extension of *Wordnet* [21]. The *SentiWordnet* is organized by synsets, the same way as *Wordnet* does. *SentiWordnet* assigns to each synset of *Wordnet* three sentiment scores: positivity, negativity, and objectivity.

Another direction to discover the hidden aspect and sentiment in reviews is to apply probabilistic approach to model the whole corpus. Griffiths and Steyvers [15] applied *LDA* to extract the hidden topics. They proposed a Markov chain Monte Carlo algorithm for inference of the model. Some other works extended basic *LDA* for improving the results. Brody and Elhadad [16] proposed a more sophisticated *LDA* model to discover the aspect hidden in reviews. A connectivity matrix is used to calculate the score, which decides the best hidden aspect number and iteration times. Then the scoring schema is used by selecting the representative words for each aspect. Titov and McDonald [17] distinguished general aspects and find-grained aspects. The model can capture ratable and global aspects which make the result more meaningful. Zhao and Jiang [18] introduced a background model and also treated general and specific aspects differently.

Multi-Aspect Sentiment Model (MAS) [12] is an extension of the previous work – MG-LDA [17], which only extract topics hidden in reviews regardless of sentiments. MAS model works in a supervised way because it requires every aspect to be rated by user. *Topic Sentiment Mixture Model(TSM)* [11] is extended from pLSA. TSM has the deficit of pLSA with inference of new documents and suffers from overfitting of the data. On the other hand, TSM does not consider the association between topic and sentiment. The words are drawn from either topic

distribution or sentiment distribution. The words are samples of a mixture of sentiment and topic, but not a combination. This makes TSM lack the ability to exact the aspect-specific opinion words. *Joint Sentiment/Topic (JST) model* [10] is a fully unsupervised model based on LDA. It can capture topic and sentiment at the same time. *Aspect Sentiment Unification model (ASUM)* [9] is a model based on JST. It is a small adaption of the JST. But it introduced the assumption that a sentence in reviews can only be referred to some an aspect and sentiment.

III. MODELS

We propose three Tag Sentiment Aspect Models (TSA) to extend the basic LDA to incorporate syntax information in different ways. In TSA1 and TSA2 models, two kinds of hidden variable are: z , aspect index, and l , sentiment label. In TSA3 model, an additional hidden variable x is introduced as an indicator besides aspect index and sentiment label. We use Gibbs sampling [15] to estimate the hidden variable. The Gibbs sampling method is a simple way to implement the inference in topic modeling, with good performance comparable with other methods and tolerant to local optimization. All the notations used here are illustrated in Table I.

TABLE I NOTATION USED IN TSA MODEL

D	the number of reviews
A	the number of aspects
S	the number of sentiments
V	the number of distinct words
N_d	the number of words in review d
T	the number of distinct tags
Θ	Multinomial distribution over aspects
Π	Multinomial distribution over sentiments
Ψ	Multinomial distribution over words
Ω	Multinomial distribution over tags
Σ	Multinomial distribution over indicators
Δ	Bernoulli distribution
λ_i	Dirichlet prior vector for σ
A	Dirichlet prior vector for θ
β_l	Dirichlet prior vector for ψ for sentiment l
Γ	Dirichlet prior vector for π
μ_l	Dirichlet prior vector for ω for sentiment l

A. Tag Sentiment Aspect Model 1(TSA1)

As the POS tag of words in reviews can be got by POS tagger, it is natural to take the POS tag of words as observed

TABLE II. TOP 10 WORDS FOR SENTI-ASPECT FOR LAPTOP DATASET

Topic Model	System and Software			Hardware and Performance			Appearance and Experience		
	Positive	Neutral	Negative	Positive	Neutral	Negative	Positive	Neutral	Negative
TSA	develop email noise couple software annoy pro con window laptop	gpu homework guess easy family year pro os pc mac	contact promies bad unacceptable stand experience refuse offer fix pay	nvidia surface cheap issu gpu button day trackpad graphic asu	i3 cell nvidia i5 chip cpu core intel graphic model	manufacture language city repute yesterday mine laptop mother board problem	generous bigger everyday people gamer wow part fact spec thing	pack release discharge lithium iron cycle charge capac life battery	desk sound button bio volume noise reason control fan compute

data in the model. The graphical presentation of TSA1 model is shown in Fig. 1(a). Tag t is generated conditioned on aspect index z and sentiment label l , along with the word w . The tag is considered as the stamp of the word. This is inspired by Wang and McCallum [19], in which the published time of the document is treated as the timestamp of words in the document.

The generative process of TSA1 model is as follows.

1. For each aspect and sentiment pair (z, l) , draw a discrete distribution over words $\psi_{z,l} \sim \text{Dir}(\beta_l)$, and a discrete distribution over tags $\omega_{z,l} \sim \text{Dir}(\mu_l)$.
2. For any a review d ,
 - a) Draw the review's sentiment distribution $\pi_d \sim \text{Dir}(\gamma)$.
 - b) For each sentiment label l , draw an aspect distribution $\theta_{dl} \sim \text{Dir}(\alpha)$.
 - c) For each word w_i and tag t_i in the review,
 - i. Choose a sentiment label $j \sim \text{Mul}(\pi_d)$
 - ii. Choose an aspect $k \sim \text{Mul}(\theta_{dj})$
 - iii. Choose word $w_i \sim \text{Mul}(\psi_{k,j})$ and tag $t_i \sim \text{Mul}(\omega_{k,j})$.

The hyper-parameters α , β , γ and μ are the pseudo-counts. It carries the prior observation of the corpus. Notice that for different sentiment label l , there are corresponding priors β_l and μ_l . That is because we use asymmetric β and μ . The asymmetric priors can exploit prior sentiment information in the corpus. For instance, elements of β corresponding to positive sentiment words should have small value for negative sentiment label, and vice versa; Elements of μ corresponding to noun tag should have large value for natural sentiment label, because the nouns often express not opinion but aspect. In TSA1 model, θ , π , ψ and ω are all the latent variables to infer. By using Gibbs sampling, we need to calculate the full conditional probabilities $P(z_i, l_i | \mathbf{z}, \mathbf{l}, \mathbf{w}, \mathbf{t}, \alpha, \beta, \gamma, \mu)$, where z_i denotes the aspect assignment for w_i , l_i denotes the sentiment assignment for w_i , \mathbf{z} denotes the aspect assignment for all word tokens except w_i , \mathbf{l} denotes the sentiment assignment for all word tokens except for w_i , and \mathbf{w} , \mathbf{t} are the word vector and tag vector for the whole corpus. During Gibbs sampling, we draw aspect and

sentiment iteratively for the word w_i according to the following probability distribution:

$$P(z_i, l_i | w_i, t_i, \mathbf{z}_{-i}, \mathbf{l}_{-i}, \alpha, \beta, \gamma, \mu) \propto \frac{\{N_{k,d}\}_{-i} + \gamma \{N_{j,k,d}\}_{-i} + \alpha}{\{N_d\}_{-i} + S\gamma \{N_{k,d}\}_{-i} + A\alpha} \frac{\{N_{w,j,k}\}_{-i} + \beta_{w_j}}{\sum_{v=1}^V \{N_{v,j,k}\}_{-i} + \beta_v} \frac{\{N_{t,j,k}\}_{-i} + \mu_{t_i}}{\sum_{t=1}^T \{N_{t,j,k}\}_{-i} + \mu_t} \quad (1)$$

where $N_{k,d}$ is the number of words assigned to sentiment label k in review d , N_d is the number of words, $N_{j,k,d}$ is the number words assigned to aspect j and sentiment k . $N_{w_i,j,k}$ is the number the word w_i assigned to aspect j and sentiment k , and $N_{t_i,j,k}$ the number the tag t_i assigned to aspect j and sentiment k . $-i$ denotes the number that excludes the i^{th} position.

Having the conditional probability, the approximate probability of θ , π , ψ and ω is estimated as follows.

$$\omega_{j,k,d} = \frac{N_{t,j,k} + \mu_{t_i}}{\sum_{i=1}^T \{N_{i,j,k}\}_{-i} + \mu_i} \quad (2)$$

$$\psi_{j,k,w} = \frac{N_{w,j,k} + \beta_w}{\sum_{v=1}^V N_{v,j,k} + \beta_v} \quad (3)$$

$$\theta_{j,k,d} = \frac{N_{j,k,d} + \alpha}{N_{k,d} + A\alpha} \quad (4)$$

$$\pi_{k,d} = \frac{N_{k,d} + \gamma}{N_d + S\gamma} \quad (5)$$

B. Tag Sentiment Aspect Model 2(TSA2)

Considering that the number of unique tag is much smaller than size of vocabulary, treating tag as stamp of words may not be proper. Additionally, as shown in Fig. 1(b), the tag is dependent on the aspect z and sentiment l , but in real situation the dependency is reverse: the aspect and sentiment are dependent on the tag of the word. When user is writing a

review, he first decides the sentiment he would like to express. If he wants to express neutral sentiment that means he just wants to give a description, he will decide to use nouns. If he wants to express an opinion that means he wants to praise or criticize something, he will decide to use adjective or adverb. Therefore, we adapt TSA1 to TSA2 as shown in Fig. 1(b). We extend the aspect distribution of document from each sentiment to each sentiment and tag pair. This simple change not only incorporates the POS tag information, but also makes the model simpler.

The generative process of TSA2 is as follows:

1. For each aspect and sentiment pair (z,l) , draw a discrete distribution over words $\psi_{z,l} \sim \text{Dir}(\beta_l)$
2. For any a review d ,
 - a) Draw the review's sentiment distribution $\pi_d \sim \text{Dir}(\gamma)$.
 - b) For each sentiment label l and tag t , draw an aspect distribution $\theta_{dl} \sim \text{Dir}(\alpha)$.
 - c) For each word w_i in the review,
 - i. Choose a sentiment label $j \sim \text{Mul}(\pi_d)$
 - ii. Choose an aspect $k \sim \text{Mul}(\theta_{dl})$, according to the word's tag t .
 - iii. Choose word $w_i \sim \text{Mul}(\psi_{k,j})$.

Like TSA1, the Gibbs sampling processing is the same. The full conditional probability is as follows.

$$P(z_i, l_i | w, t, z_{-i}, l_{-i}, \alpha, \beta, \gamma, \mu) \propto \frac{\{N_{k,d}\}_{-i} + \gamma}{\{N_d\}_{-i} + S\gamma} \frac{\{N_{j,kt,d}\}_{-i} + \alpha}{\sum_{v=1}^V \{N_{kt,d}\}_{-i} + A\alpha} \frac{\{N_{w_i,j,k}\}_{-i} + \beta_{w_i}}{\sum_{v=1}^V \{N_{v,j,k}\}_{-i} + \beta_v} \quad (6)$$

where the major difference with TSA1 is that the 4th part of the TSA1's conditional probability is disappear and the 2nd part is different on the subscript. In TSA2, the times of words in review d assigned to aspect j and sentiment k is counted on every type of tag. The approximate probability of θ, π, ψ and ω is able to estimate as follows.

$$\psi_{j,k,w} = \frac{N_{w,j,k} + \beta_w}{\sum_{v=1}^V N_{v,j,k} + \beta_v} \quad (7)$$

$$\theta_{j,kt,d} = \frac{N_{j,kt,d} + \alpha}{N_{kt,d} + A\alpha} \quad (8)$$

$$\pi_{k,d} = \frac{N_{k,d} + \gamma}{N_d + S\gamma} \quad (9)$$

C. Tag Sentiment Aspect Model 3(TSA3)

There is a deficit in above TSA models. The aspect distribution θ is extended by T types of tag. This is based that the tag of words indicates whether the word is about aspect or opinion. We draw a different θ exactly according to the type of the tag. This is not proper because it implies

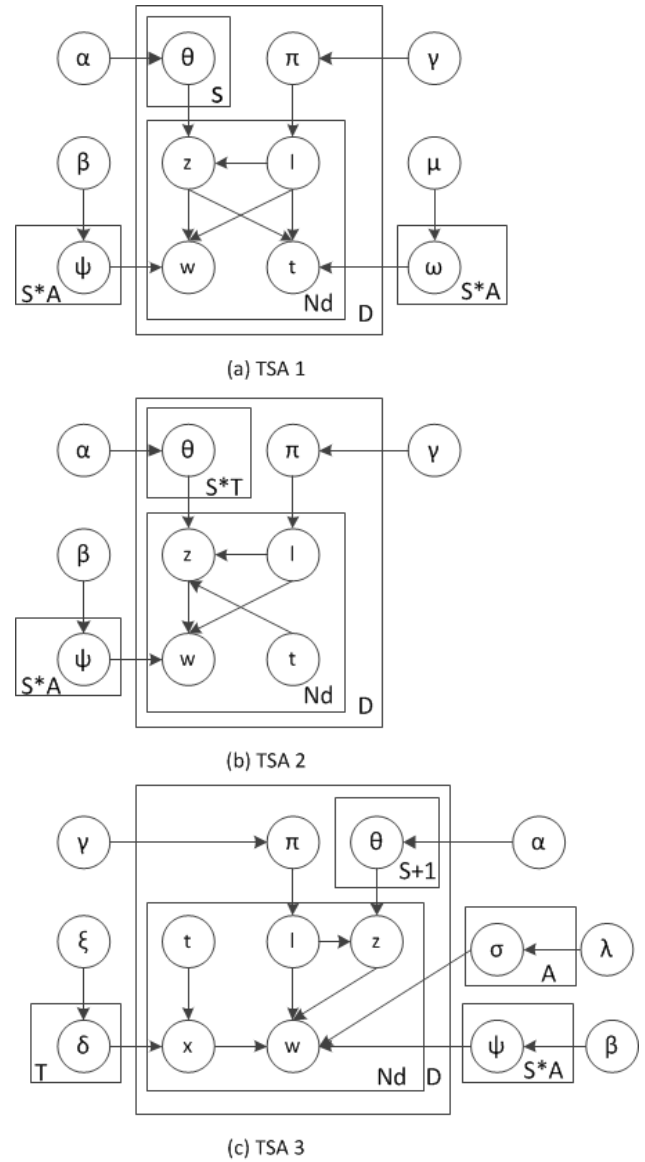


Figure 1. Tag Sentiment Aspect Models

that we applied a strict rule that a type of tag always serve as the same function. In fact, there often be exceptions. For example, the noun 'problem' usually serves as an opinion word. To overcome this deficit, in TSA 3, we introduce a indicator variable x to indicate whether the word is serving as an aspect word or an opinion word. If x equals 0, the word serves as an aspect word and it will be drawn by distribution σ , which is a distribution set related to A aspects. If x equals 1, the word serves as an opinion word and it will be drawn by distribution ψ , a distribution set related to A aspects and S sentiments. The variable x is drawn from distribution δ . Different type of tag has a different δ . TSA3 is shown in Fig. 1(c). Notice that there is $(S+1)$ θ for each review because the extra one θ is for the aspect word.

The generative process of TSA3 is as follows.

1. For each aspect and sentiment pair (z,l) , draw a discrete distribution over words $\psi_{z,l} \sim \text{Dir}(\beta_l)$, for

every aspect z , draw discrete distribution over words $\sigma_z \sim \text{Dir}(\lambda)$. For every tag t , draw Bernoulli distribution $\delta \sim \text{Dir}(\xi)$

2. For any a review d ,
 - a) Draw the review's sentiment distribution $\pi_d \sim \text{Dir}(\gamma)$.
 - b) For each sentiment label l , draw an aspect distribution $\theta_{dl} \sim \text{Dir}(\alpha)$, and an extra aspect distribution $\theta_d \sim \text{Dir}(\alpha)$.
 - c) For each word w_i in the review,
 - i. Choose indicator $x \sim \text{Ber}(\delta_i)$, according to its tag t
 - ii. If x is 0, choose an aspect $k \sim \text{Mul}(\theta_{dl})$ and choose word $w_i \sim \text{Mul}(\sigma_k)$
 - iii. If x is 1, choose a sentiment label $j \sim \text{Mul}(\pi_d)$, choose an aspect $k \sim \text{Mul}(\theta_{dj})$ and choose word $w_i \sim \text{Mul}(\psi_{kj})$.

The full conditional probability is as follows.

$$P(z_t = j, l_t = k | w, t, z_{-t}, l_{-t}, \alpha, \beta, \gamma, \mu) \propto \frac{\{N_{x,t}\}_{-i} + \xi_{x_t}}{\{N_d\}_{-i} + \xi_m}$$

$$x=1 \quad \frac{\{N_{w_i,j,k}\}_{-i} + \beta_{w_i}}{\sum_{v=1}^V \{N_{v,j,k}\}_{-i} + \beta_v} \frac{\{N_{k,d}\}_{-i} + \gamma}{\sum_{l=1}^S \{N_{l,d}\}_{-i} + \gamma} \frac{\{N_{j,k,d}\}_{-i} + \alpha}{\{N_{k,d}\}_{-i} + A\alpha} \quad (10)$$

$$x=0 \quad \frac{\{N_{w_i,j}\}_{-i} + \lambda_{w_i}}{\sum_{v=1}^V \{N_{v,j}\}_{-i} + \lambda_v} \frac{\{N_{j,d}\}_{-i} + \alpha}{\sum_{z=1}^A \{N_{z,d}\}_{-i} + \alpha} \quad (11)$$

In TSA3, ξ_t is asymmetric and is incorporated with prior information. The approximate probability of δ , θ , π , ψ and σ is able to estimate as follows.

$$\delta_{x,t} = \frac{\{N_{x,t}\}_{-i} + \xi_x}{\sum_{m=0}^1 \{N_{m,t}\}_{-i} + \xi_m} \quad (12)$$

$$\theta_{j,k,d} = \frac{N_{j,k,d} + \alpha}{N_{k,d} + A\alpha} \quad (13)$$

$$\theta_{j,d} = \frac{N_{j,d} + \alpha}{N_d + A\alpha} \quad (14)$$

$$\psi_{j,k,w} = \frac{N_{w_i,j,k} + \beta_w}{\sum_{v=1}^V N_{v,j,k} + \beta_v} \quad (15)$$

$$\sigma_{j,w} = \frac{N_{w_i,j} + \lambda_w}{\sum_{v=1}^V N_{v,j} + \lambda_v} \quad (16)$$

$$\pi_{k,d} = \frac{N_{k,d} + \gamma}{N_d + S\gamma} \quad (17)$$

IV. EXPERIMENTAL SETUP

We use the dataset of electronic device reviews from Jo and Oh [9]. We investigate the power of our models and select the Laptop and DigitalSLR categories to form our

experimental dataset. We compare our TSA models by the power of sentiment classification of the review, and the power of discovering latent aspect and extracting the aspect-specific sentiment words. We also give a comparison between our models and existing models JST [10] and ASUM [9].

A. Preprocessing

For each dataset, we choose 1000 reviews including 500 positive and 500 negative. The original reviews are rated by 5-star system. We discard the 3 star reviews, and treats 1 or 2 star reviews as negative ones, 4 or 5 star reviews as positive ones. Then preprocessing is performed on DigitalSLR and Laptop dataset. NLTK (Natural Language Toolkit) [22], a software package implemented by PYTHON is used for preprocessing. First of all, the sentence segmentation algorithm is performed to obtain the sentences of each review. A POS tagger is then tagging every sentence. Afterwards, we remove the punctuation, numbers, and non-alphabet tokens. We filter out the tokens using a stop-word list [23]. For integrating syntax information, we consider 'NN', 'JJ', 'VB', 'RB', the 4 types of tag and ignore the others for the simplicity. After preprocessing, every corpus contains 1000 reviews. The DigitalSLR dataset has 83931 words with 5657 distinct words and the Laptop dataset has 81648 words with 5318 distinct words.

B. Prior Information

There are two key elements for incorporating prior information in TSA models. One key is carefully tuned hyper-parameters, the other is the initialization of Gibbs sampling. In the experiment, we use asymmetric hyper-parameters β to exploit the sentiment bias. We use a positive word list and a negative word list. If a word is in the positive list, the corresponding value of the β will be large for positive sentiment aspect and small negative sentiment, and vice versa. The exactly value to set is according to the actual experiment, which will be described in the following section. Considering the sentiment word list, we use the paradigm word list used in [9]. The sentiment words are applied in the initialization of Gibbs sampling. The word token in the sentiment word list is assigned to the corresponding sentiment label. During the iterative sampling process, the initialization effect becomes weak, so the iteration times should not be too large. We empirically set the iteration times to be 1000. The POS tag information is incorporated in the same way as the sentiment words list. In the initialization, the sentiment label of the word with 'NN' tag is drawn with distribution whose probability is large on the neutral label, and the label of the word with 'JJ, VB, RB' tag is drawn with distribution whose probability is large on the positive/negative label. Asymmetric prior μ_l is used for different sentiment label in TSA 1, and asymmetric prior ξ_t is used for different type of tag respectively in TSA 3.

C. Tasks

Two abilities of our models are investigated. The first is the ability to discover latent aspect word and aspect specific opinion word. The result of TSA models is analyzed to evaluate this ability. We compare our TSA models with existing approaches. The second ability is the power of sentiment classification. The sentiment distribution π , which indicates the proportion of each sentiment in the review, can achieve the task of sentiment classification. If the positive sentiment has a higher probability than the negative sentiment, the review is classified as positive review, and vice versa. We compare our models with ASUM, JST and SVM with different features.

V. EXPERIMENTAL RESULTS

We first examine the automatically discovered aspects and senti-aspects from reviews by our TSA models. Then we examine the sentiment classification results.

A. Aspect and Senti-Aspect Discovery

To train our TSA models, we first set the parameters used in the model. The number of aspect is set to be 50, the prior α is set to be 1.0, according to previous works which show α should be set to 50 for topics. A symmetric prior γ is used that we assume no prior knowledge of the sentiment distribution. The value is set to 1, which means all sentiment probabilities are equal. As mentioned above, prior β should be tuned carefully for its key effect to incorporate the priors. In TSA1 and TSA2, for positive aspect, we set elements of β vector to be 0 for negative words and other elements to be 0.01. For negative aspect, we set elements of β vector to be 0 for positive words and other elements to be 0.01. For neutral aspect, we use symmetric β set to be 0.01. In TSA1, the prior μ is set in the same way: 5 for elements corresponding to 'NN' and 1 for others for neutral aspect and 5 for elements corresponding to 'JJ', 'RB', 'VB' and 1 for others for positive/negative aspect. In TSA3, asymmetric ξ is used by setting 5 for 'NN' and 1 for 'JJ', 'RB', 'VB' for x is 0, and 5 for 'JJ', 'RB', 'VB' and 1 for 'NN' for x is 1. Notice that, there are 3 sentiment labels in TSA1 and TSA2, and 2 sentiment labels in TSA3. So the β prior of TSA3 is setting as the β for positive and negative set is TSA1 or TSA2, and the new λ prior is set as the neutral one in TSA1, a symmetric prior. All these settings are used when there is prior of sentiment word list. In a random initialization context, we ignore these asymmetric priors. Instead, we use symmetric priors.

As the output of the model is the word distributions, which is also called language model. One language model presents how frequent the word will occur under certain aspect or aspect/sentiment pair. For TSA1 and TSA2, 3 (sentiment label) * 50 (aspect) word distribution will be obtained. For convenience of analysis, we place the extra 50 distributions with a virtual neutral label. The 50 distributions in TSA3 denotes only aspect, and neutral distribution in TSA1 and TSA2 denotes aspect and opinion with neutral sentiment

label. We show the results in Table II. We select 3 aspects out of 50 for each model, and for every distribution we examine the top 10 words.

In Table II, we list three aspects drawn from the TSA models, and the labels of aspects are annotated manually. For 'system and software', the top words 'xp', 'vista', 'mac', which are different names of operating systems, are contained. The word 'program', 'windows', 'os', which are the concept of the 'system', are also contained. For another aspect, 'Hardware and performance', we get the words 'i3', 'i5', 'i7', which indicate a special architecture of CPU, and we also get the words 'cpu', 'intel', 'chip' which explicitly indicate the 'hardware' aspect. The third aspect is 'appearance and experience', the appearance covers the style of laptop, screen, color, and weight, etc. The experience is about the joyment of customer to use this laptop. The second criteria requires little overlap between different aspects. It is obviously shown in Table II.

TSA models discover senti-aspect as well. Under each aspect, two distributions corresponding to positive and negative sentiment are also inferred. For 'system and software', 'happy', 'best', 'nice', 'easy', and 'safe' are top words when talking about the positive side of the aspect, 'bad', 'refuse', 'slow', 'hard', 'noise', and 'lose' are top words when talking about the negative side. For 'hardware and performance' aspect, when describing positive side, 'worth', 'high', 'better', and 'latest' are often used, and when describing negative side, 'hot', 'problem', 'noise', 'drop', 'bad' are often used in reviews. For 'appearance and experience', 'enjoy', 'bigger', and 'sonystyle' are used to express positive attitude and 'claim', 'provide', 'fail' and 'cancel' are used to express negative attitude. It can be noticed that the overlap of senti-aspect is high because there are two types of sentiment words. One is common sentiment words like 'good', 'bad', 'hate', and 'love', etc. The other is specific to a certain aspect, like 'hot' is positive for appearance, but negative for performance.

The power of three TSA models is different. For TSA1, the tags of words are treated as observed data. An extra distribution ω is introduced to reflect the tag informations. An intuition thought is that the word distribution ψ affected little by tag information, for this information almostly is coded in distribution ω . Therefore, the top words produced by ψ may not have high correlation with tags of words. This is shown in Table II. The top words of senti-aspect by TSA1 are the mixture of different tag types. For 'system and software', the positive aspect has words with noun tag like 'develop', 'email' beside words with adjective tag like 'couple', 'pro'. For 'appearance and experience' aspect, it is similar. The negative aspect has words with adjective tag, like 'noise', 'bio', besides words with noun tag, 'reason', 'volumn'. For TSA2, the tag information is integrated by the document-aspect distribution, θ . The intuition is that the top words from distribution ψ will have high correlation with tags. But another problem is that as we use different θ according to the tag to draw a word, which means we use

the strict rule of that, words with different tag are always play different roles. So, the TSA2 may suffer with the loss of ability with exploiting nouns used in positive or negative aspect. The intuition is verified again as shown in Table III. For both positive and negative aspect, words with noun tag are hardly seen. For ‘system and software’, the top words in the positive aspect are all adjective except for the word ‘3d’, and the top words in the negative aspect are all adjective except for the verb ‘consist’ and the noun ‘drive’. Similar situation is also observed in ‘hardware and performance’ aspect and ‘appearance and experience’ aspect. TSA3 is designed to integrate tag information to ψ distribution but not to loss flexibility and its ability to explore noun words in sentimental aspects. In TSA3, the indicator variable indicates that a word is the aspect word or the sentiment word under an aspect. The variable x is drawn from Bernoulli distribution conditioned on tag of the word. For adjective, verb and adverb, x is highly probable to be 1, indicating a sentiment word under an aspect, while noun is highly probable to be zero, indicating an pure aspect word. By introducing the Bernoulli distributions, the rule is relaxed with the nouns as sentiment words in the low probability. In Table III, the nouns such as ‘damage’ and ‘waste’ are presented as top words in ‘hardware and performance’ negative aspect, and the nouns such as ‘virus’, ‘problem’ and ‘lose’ are presented as top words in ‘system and software’ negative aspect. For positive aspects, the nouns such as ‘power’, ‘monitor’, ‘design’ and ‘budget’ are also appeared as sentiment words.

The same situation is also observed when applying TSA models to the DigitalSLR dataset, we list some aspects and sentiment words in Table III.

TABLE III. ASPECT WORDS AND CORRESPONDING SENTIMENT WORDS FOR DIGITALSLR DATASET LIST

Aspect Words	Sentiment Words
compact depth picture color olympus eye iso electron	small long light higher manual fun weather love support heavier stain dust waste disappoint shallow horrible worry
camera focus len shoot process zoom angle digit	long smooth easy original high wide motor happier short expensive problem loss heavier difficult change claim
strobe battery d700 handbrake aim camcorder neon viewfinder	profession bright long light adjust detail friend improve amateur spend refund defect stain useless recharge lost

B. Sentiment Classification

We use distribution π for sentiment classification task on the review level. As mentioned above, we discard neutral reviews and do the evaluation only on positive and negative ones. The distribution $\pi_{k,d}$ presents the probability of sentiment k in review d , we compare the probability of sentiment positive and negative, and assign larger label to the review. We compare TSA models with existing approaches. The result is shown in Table IV. We also

exploit the effect of priors, and the result shows the prior enhance the performance largely.

TABLE IV. SENTIMENT CLASSIFICATION ON DIGITALSLR AND LAPTOP DATASETS WITH Laptop TSA MODEL AND PREVIOUS APPROACHES, AND THE PLUS MEANS WITH PRIOR INFORMATION INCORPORATED

	Laptop Dataset			DigitalSLR Dataset		
	pos	neg	overall	pos	neg	overall
TSA1	58.8%	53.2%	56%	60.8%	56%	58.4%
TSA2	50.8%	54.6%	53.7%	59%	53.6%	56.3%
TSA3	59%	53.4%	56.2%	61%	58.8%	59.9%
TSA1+	86.6%	82.4%	85%	83.4%	86.2%	83.8%
TSA2+	80.8%	85.4%	83.1%	82.8%	82.4%	82.6%
TSA3+	84.8%	87.4%	86.2%	86.8%	84.4%	85.6%
ASUM	58%	52.2%	55.1%	54.8%	56.2%	55.5%
ASUM+	85.8%	84.6%	85.2%	86.4%	82.8%	84.6%
JST	54%	52.8%	53.4%	50.6%	55.2%	52.9%
JST+	78.6%	84.4%	81.5%	82.8%	76.6%	79.7%

In Table IV, first observation is that the prior information could enhance the accuracy of sentiment classification. The average accuracy is 55% and the prior improves that to about 80%. TSA2 performs worse than TSA1 and TSA3. That is because in TSA2, the strict rule that words with different tag plays the different roles in introducing the noise. TSA1 and TSA3 almost have the same power on sentiment classification, although the power of TSA3 on aspect discovery is better than TSA1. That is because the distribution π incorporates the information of ω but the ψ distribution does not. We also compare our models with JST [10] and ASUM [9]. JST model is worse than our TSA models. ASUM is sometimes better than TSA1 and TSA2, but not better than TSA3.

VI CONCLUSION AND FUTURE WORK

The ‘‘bag of words’’ assumption is suitable for traditional text classification but when comes to opinion mining, it is not good one. Opinion is expressed in the more complicated way. We need to explore more information hidden in the natural language. The tag of word is a good attempt. Our TSA models are to incorporate this kind of information. The results of our approach have better effectiveness as shown in Table IV.

The future works have two directions: one is to exploit other language information into the model. For example, the dependency relation can be used. A synonym thesaurus can be used to explore the relations between aspect words. The other direction is to design the prior information better and adjust the TSA model to fit the corpus. We also can add an additional background language model to capture the

common across the language. In the TSA models, the prior probabilities of different tags are still fixed manually. We could further use the unsupervised machine learning method to train a model to determine these probabilities.

REFERENCES

- [1] S. Blair-Goldensohn, K. Hannan, R. McDonald, T. Neylon, G. A. Reis, and J. Reynar, "Building a Sentiment Summarizer for Local Service Reviews," In WWW Workshop on NLP in the Information Explosion Era (NLPiX), USA, 2008.
- [2] 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, USA, 2004, pages 168 – 177.
- [3] 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, USA, 2005, pages 339 – 346.
- [4] 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, USA, 2002, pages 79 - 86.
- [5] 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, USA, 2004.
- [6] P. D. Turney, "Thumbs Up or Thumbs Down?: Semantic Orientation Applied to Unsupervised Classification of Reviews," In Proceedings of the 40th Annual Meeting on Association for Computational Linguistics, USA, 2001, pages 417–424.
- [7] P. D. Turney and M. L. Littman, "Unsupervised Learning of Semantic Orientation from a Hundred-billion-word Corpus," CoRR, cs.LG/0212012, 2002.
- [8] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent Dirichlet Allocation," Journal of Machine Learning Research, Vol.3, 2003, pages 993-1022.
- [9] Y. Jo and A. Oh, "Aspect and Sentiment Unification Model for Online Review Analysis," In Proceedings of WSDM, 2011
- [10] C. Lin and Y. He, "Joint Sentiment/Topic Model for Sentiment Analysis," In Proceeding of the 18th ACM Conference on Information and Knowledge Management, USA, 2009, pages 375 – 384.
- [11] Q. Mei, X. Ling, M. Wondra, H. Su, and C. Zhai, "Topic Sentiment Mixture: Modeling Facets and Opinions in Weblogs. In Proceedings of the 16th International Conference on World Wide Web, USA, 2007, pages 171-180.
- [12] I. Titov and R. McDonald, "A Joint Model of Text and Aspect Ratings for Sentiment Summarization. In Proceedings of International Conference on ACL, 2008, pages 308-316.
- [13] T. Griffiths, M. Steyvers, D. Blei, and J. Tenenbaum, "Integrating Topics and Syntax," In Advances in Neural Information Processing Systems 17, 2004.
- [14] Y. Lu, M. Castellanos, U. Dayal and C. Zhai, "Automatic Construction of a Context-Aware Sentiment Lexicon: An Optimization Approach," In Proceedings of WWW, 2011.
- [15] T. Griffiths and M. Steyvers, "Finding Scientific Topics," In Proceedings of the National Academy of Sciences (101), 2004, pages 5228–5235.
- [16] S. Brody and N. Elhadad, "An Unsupervised Aspect-Sentiment model for Online Reviews," In Proceedings of the 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics, USA, 2010, pages 804 - 812.
- [17] I. Titov and R. McDonald, "Modeling Online Reviews with Multi-grain Topic Models," In the Proceeding of the 17th International Conference on World Wide Web, USA, 2008, pages 111–120.
- [18] X. Zhao, J. Jiang, H. Yan, and X. Li, "Jointly Modeling Aspects and Opinions with a MaxEnt-LDA Hybrid," In Proceedings of the International Conference on Empirical Methods in Natural Language Processing, USA, 2010, pages 56 – 65.
- [19] X. Wang, A. McCallum, "Topics over Time: a Non-Markov Continuous-Time Model of Topical Trends," In Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, USA, 2006.
- [20] SentiWordNet, <http://sentiwordnet.isti.cnr.it> [retrieved: Aug. 2013]
- [21] WordNet, <http://wordnet.princeton.edu> [retrieved: May 2013]
- [22] NLTK, Natural Language Toolkit, <http://www.nltk.org> [retrieved: Aug. 2013]
- [23] <http://www.lextek.com/manuals/onix/stopwords1.html> [retrieved: May 2013]