Sentiment Topic Model with Decomposed Prior

Chengtao Li* Jianwen Zhang †‡ Jian-Tao Sun† Zheng Chen†

Abstract

This paper deals with the problem of jointly mining topics, sentiments, and the association between them from online reviews in an unsupervised way. Previous methods often treat a sentiment as a special topic and assume a word is generated from a flat mixture of topics, where the discriminative performance of sentiment analysis is not satisfied. A key reason is that providing rich priors on the polarity of a word for the flat mixture is difficult as the polarity often depends on the topic. To solve the problem we propose a novel model. We decompose the generative process of a word's sentiment polarity to a two-level hierarchy: the first level determines whether a word is used as a sentiment word or just an ordinary topic word, and the second level (if the word is used as a sentiment word) determines the polarity of it. With the decomposition, we provide separate prior for the the first level to encourage the discrimination between sentiment words and ordinary topic words. This prior is relatively easy to obtain compared to the concrete prior of the word polarities. We construct the prior based on part-of-speech tags of words and embed the prior into the model. Experiments on four real online review data sets show that our model consistently outperforms previous methods in the task of sentiment analysis, and simultaneously performs well in the sub-tasks of discovering ordinary topics, sentiment-specific topics, and extracting topic-specific sentiment words.

1 Introduction

Nowadays, at all kinds of web sites, users are never bored with writing an overwhelming amount of reviews on products/services, books/movies, events, celebrities, etc. Analyzing online reviews is an important job for many people. According to reviews, consumers make the purchase decision, producers receive valuable feedbacks and organizations analyze popular opinions [17]. For someone analyzing the reviews, the key information she wants to obtain is: (1) The overall opinion/sentiment of the review, e.g., good/bad, like/dislike, support/oppose. (2) The topics (or aspects/features/parts) of the target (e.g., a product) talked about in the reviews. For example, reviews on a laptop often contain detail comments on its different topics such as "CPU performance", "battery life", "display quality", etc. (3) The association between topics and sentiments, including the topics associated with a specific sentiment and the sentiments associated with a specific topic. For example, what features of a laptop receive positive sentiment? And what opinions are expressed on the topic "CPU performance", fast, hot, noisy or something else? It is a heavy job for a human being to get the above information by reading the reviews one by one. For one thing, the amount of reviews is often huge. For another, the topics and sentiments are statistical information not easily captured by a human being. As a result, it is of great value to automatically analyze the reviews to extract topics, sentiments and the associations between them. This problem received surging attention in both academic and industry recently [15, 19, 13, 5, 12, 10, 16, 14, 17].

The problem is different from traditional *sentiment classification* [18], where only the overall sentiment of a review (i.e., document-level sentiment) is cared about. In this problem, we also care about topic-level and word-level sentiments. Actually the problem consists of three sub tasks. First, sentiment analysis, to judge the document-level sentiment of a review; second, named *senti-topic* and *topic* extraction (Tabel 1), to extract different topics associated with a specific sentiment or no sentiment; third, named *topic-senti word* extraction, to extract the sentiment words associated with a specific topic, i.e., to discover how opinions with various sentiment polarities are expressed by elementary words in a specific topic. The third is not only a mandatory task but also a necessary approach to help sentiment analysis on topic level and document level. The reason is that many sentiment expressions are topic-specific. For example, the word "fast" describing the battery consuming often expresses a negative sentiment, but the polarity of the same word describing the CPU performance is positive. Then it is of hope to improve the sentiment analysis by recognizing the topic and using a sentiment model which is topic-specific rather than topic-independent. Thus the third sub task paves the road of the previous two sub tasks.

It is not easy to well solve the problem. Previously there were some complex models proposed to jointly model topics and sentiments based on topic models, however, the following two fundamental challenges have not been conquered.

First, how to represent and couple the topics and sentiments. Representing topics is relative easier thanks to the extensive studies on topic models. In contrast, the linguistic mechanisms of sentiment expression, especially embedded in specific topics, are of fewer studies. The widely adopt-

[*] The Institute for Theoretical Computer Science(ITCS), Institute for Interdisciplinary Information Sciences, Tsinghua University. This work was supported in part by the National Basic Research Program of China Grant 2011CBA00300, 2011CBA00301, the National Natural Science Foundation of China Grant 61033001, 61061130540, 61073174. lichengtao2010@gmail.com

[†] Microsoft Research Asia. {jiazhan,jtsun,zhengc}@microsoft.com

[‡] The corresponding author.

ed approach is to treat each sentiment polarity as a special topic in a topic model and infer them together with the ordinary topics [13, 14, 10, 12]. This approach is effective in topic extraction. However, as the number of topics is often much larger than the number of sentiment polarities, without special mechanisms that encourage discrimination between sentiments and ordinary topics, it tends to bias to topic extraction and harm the performance of sentiment analysis, as illustrated in the experiments of Section 5.

Second, in an unsupervised model, how to distinguish a senti-topic from an ordinary topic and make sure the model sentiments are accordant with the human sentiments. The paradox comes from that an ideal approach in practice is expected to be unsupervised while sentiment analysis is a discriminative problem. For shopping reviews there is sometimes a rating as the label for the overall sentiment of a review, however, many other types of reviews have no such labels. Besides, in reviews there are generally not any supervision information at the topic level. Consequently most of the recent works are committed to unsupervised methods [13, 5, 10, 12]. But in essence sentiment analysis is a discriminative task because its sentiment taxonomy should be accordant with that of human beings. As a result, the key issue is how to incorporate in proper prior on the correspondence between the model sentiments and the "true" human sentiments. One popular approach is to provide some polarity lexicons such as general polarity words ("good", "bad", etc.) for each model polarity, to shape the prior word distributions of a model polarity towards the corresponding polarity of human being. However, this approach has its limitations. If the polarity lexicons are not rich, containing only the general polarity words like "good" and "bad", the impact of the prior is very limited because they are of little help to identifying the topic-specific sentiment words, which are the majority. Conversely, requiring extensive topic-specific polarity lexicons as prior is a cycling problem: it is right the target of the third sub task (topic-senti words extraction) of this paper. As a result, we need to seek for other approaches.

In this paper we propose a new senti-topic model to conquer the above challenges. In the model, instead of adopting a flat mixture, we decompose the generative process of a word's sentiment polarity to a hierarchy of two levels: the first level determines whether the word is used as a sentiment word or just an ordinary topic word, and the second level (if the word is used as a sentiment word) determines which polarity it is. With this decomposition, we provide separate prior to the first phase to encourage discrimination between sentiment words and ordinary topic words, which is expected to help the discrimination between different sentiment polarities at the second level in turn. Compared to the concrete prior of which polarity a word is, the prior of whether a word is a sentiment word is relative easier to obtain. For example, some researchers observed that whether a word functions as a sentiment word is closely related with its linguistic category (the part-of-speech (POS)) [20, 3, 6, 2]. According to this observation, we construct the prior based on POS analysis and embed the prior into the model. We perform experiments on four real online review data sets, and find the proposed model is effective on discovering senti-topics and extracting topic-specific sentiment words. More importantly, compared to several state-of-art models, the proposed model consistently achieves much better accuracies on document level sentiment prediction.

2 Related Work

Closely related to this paper, there are some previous works attempting to jointly model topics and sentiments, as briefly introduced as follows.

Joint Sentiment-Topic model (JST) [13] is a two-level senti-topic model based on Latent Dirichlet Allocation (L-DA) [4]. In JST a sentiment is represented as a topic and a word is assumed to be generated from a flat mixture of all the topics (including ordinary topics and senti-topics). JST assumes every word has a sentiment polarity, which is not the case since some words do not express any sentiment and thus should have no influence on the result of sentiment analysis.

Reverse Joint Sentiment-Topic Model (RJST) [14] reconstructs the JST model from a dual viewpoint on the association between topics and sentiments. JST models sentiment specific topics, i.e., topics are generated conditioned on a sentiment polarity, while in RJST the association direction is reversed: sentiments are generated conditioned on a topic. In other words, RJST models topic level sentiments. To obtain the document level sentiment, it has to add an extra level of variables to link the topic sentiments and the document sentiment. The author also observed that its performance on document level sentiment analysis is worse than that of JST.

Aspect and Sentiment Unification Model (ASUM) [10] is similar to JST. The only difference is that, it models each sentence as a unit of a senti-topic, while JST (and our proposed model) models each word as the unit. The assumption that all words from a sentence expresses the identical topic and sentiment is sometimes too strong. It may work well on long documents with many sentences. However, it is not applicable for short reviews, which is also verified in Section 5.

All of the three models mentioned above model sentiments as special topics, and provide sentiment seed words in order to encourage model sentiments accordant with human sentiments and improve the performance of sentiment analysis. The limitations of this approach is pointed out in Section 1 (the second challenge). Without further priors indicating whether a word is sentiment word or not to encourage the differentiation between ordinary topics and sentiments, the senti-topics tend to be contaminated by non-senti-topics, and it subsequently harms the performance of sentiment analysis. We overcome the drawback in this paper by decomposing the prior, where the separate prior on whether a word is senti-topic is incorporated in.

There are some other models that jointly model topics and sentiments but used for other tasks. Topic Sentiment Mixture model (TSM) [15] is proposed to analyze topic life

Table 1: Terminologies

Senti-topic	A multinomial distribution over words that presents a pair of topic and sentiment.
Topic-senti words	Words used to express sentiment in a specific topic. For example, the word “fast” in topic “CPU performance” expresses a positive polarity. The same word in the topic “battery life” expresses a negative polarity.
Topic-non-senti words	Words under a specific topic that are not used to express sentiments, for example, the word “frequency” under topic “CPU performance” is not used to express sentiments.
Part-of-speech group	We categorize different kinds of part-of-speeches into groups. A group is called a part-of-speech group, the part-of-speeches of which have similar likelihood to express sentiments. The specific groups used in this paper are in Tabel 5.

cycles and sentiment dynamics in a corpus. It jointly models topics and sentiments in the corpus based on the Probabilistic Latent Semantic Indexing (PLSI) model [9]. Several drawbacks of the model are pointed out in previous literatures [13, 10, 12]. For example, the topic and sentiment components are separately modeled without association, hence it cannot solve the problem in this paper. FACTS, CFACTS, FACTS-R and CFACTS-R [12] utilize the sequence model of HMM-LDA [8] to capture the sequential dependency patterns between adjacent words in a sentence when expressing sentiments. In addition, they target at pre-specified topics, requiring the category labels for the topics. CTRF [21] is proposed to incorporate more flexible priors for topic model, but it’s not clear how to perform the document-level sentiment prediction, which makes it different from our work.

There are some other works studying the contribution of POS tags of words to sentiment analysis. In[20], base noun phrases such as *NN, JJ* are used as candidate feature terms to extract sentiments. In [3], complex adjective phrases are used to discover opinion propositions. Authors of [2] propose a way to score the strength of sentiment expressed by adverb-adjective combinations like “very good”, and use the score to perform sentiment analysis on topic level. All these works suggest that POS tags of words have strong correlations with whether words are used to express sentiments. That’s one of the reasons why we incorporate the part-of-speech tags of words as priors in our model, as shall be discussed later.

3 Model

In this section we introduce the proposed generative model, Senti-Topic model with Decomposed Prior (STDP).

3.1 Generative Process. STDP simulates the the generative process of writing a review as the process shown in the following scenario:

1. A user decides to write a review to express a distribution of *sentiments*. In each sentiment she decides a distribution of *topics* that she is going to comment on.
2. She first picks up a *part-of-speech group* (e.g., Verb related, see Tables 1&5) of the word she will use, following syntax rules.
3. According to the part-of-speech group of the word, she makes a choice over expressing sentiments or talking about ordinary topics.
4. If she decides to express a sentiment, she selects a senti-topic associated with that sentiment. Otherwise, she selects an ordinary topic without any sentiment polarities.

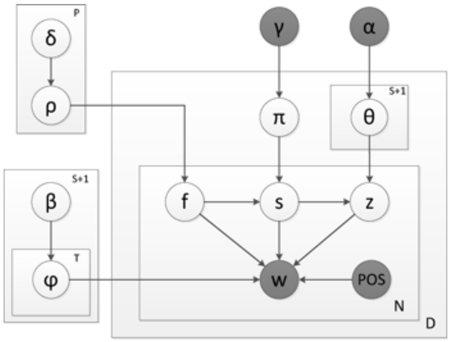

Figure 1: Graphical representation of STDP Model. Shaded nodes are observed.

5. Finally she selects a word from the topic and writes down the word. Then she goes back to step 2 for the next word.

Suppose that there are S sentiment polarities indexed by $\mathcal{S} = \{1, \ldots, S\}$. For each sentiment s, there are T senti-topics associated with it, indexed by $\mathcal{T} = \{1, \ldots, T\}$. In addition, there are another T ordinary topics which do not carry any sentiments, i.e., are associated with a *non-sentiment*. For convenience of notations, we use $S + 1$ to index this “non-sentiment”. Each topic, either senti-topic or ordinary topic, is a distribution over words, denoted as $\boldsymbol{\phi}_{st}$, $s \in \mathcal{S}_+ = \mathcal{S} \cup \{S + 1\}$. The dimension of $\boldsymbol{\phi}_{st}$ is W, the vocabulary size. Then there are totally $(S + 1) \times T$ topics, $S \times T$ for senti-topics, T for ordinary topics. Consider a review corpus with D documents, the formal generative process is as follows, corresponding to the probabilistic graphical model depicted in Figure 1:

1. **Senti-topics generation**. For every pair of topic $t \in \mathcal{T} = \{1, \ldots, T\}$ and sentiment $s \in \mathcal{S}$, draw a senti-topic word distribution $\boldsymbol{\phi}_{ts} \sim Dirichlet\left(\boldsymbol{\beta}_{ts}\right)$.
2. **Ordinary topics generation**. For every topic t, draw an ordinary topic word distribution $\boldsymbol{\phi}_{t(S+1)} \sim Dirichlet(\boldsymbol{\beta}_{t(S+1)})$.
3. **Sentiment prior generation**. For each group of part-of-speech p, draw a distribution $\boldsymbol{\rho}_p \sim Beta(\boldsymbol{\delta}_p)$. The distribution $\boldsymbol{\rho}_p$ describes how likely words in group p would be used to express sentiments.
4. **Document generation**. For each document d

(a) Draw a S-dimensional sentiment distribution $\boldsymbol{\pi}_d$ on S sentiment polarities: $\boldsymbol{\pi}_d \sim Dirichlet(\boldsymbol{\gamma})$. $\boldsymbol{\pi}_d$ indicates the document’s overall tendency on sentiment polarities.

(b) For each $s \in \mathcal{S}_+$, draw a T-dimensional topic distribution $\boldsymbol{\theta}_{ds}$ on T topics: $\boldsymbol{\theta}_{ds} \sim Dirichlet(\boldsymbol{\alpha})$. $\boldsymbol{\theta}_{ds}$ indicates the document's overall tendency on the topics associated with s.

(c) **Word generation**. For each word w in document d (with the part-of-speech group p observed):

i) Draw the flag $f \sim Bernoulli(\boldsymbol{\rho}_p)$ to indicate whether w functions as a sentiment word.

If $f = true$ (w functions as a sentiment word):

ii) Draw a polarity $s \sim Multinomial(\boldsymbol{\pi}_d)$

iii) Draw a senti-topic t associated with the polarity: $t \sim Multinomial(\boldsymbol{\theta}_{ds})$

iv) Draw a word from the senti-topic: $w \sim Multinomial(\boldsymbol{\phi}_{ts})$.

If $f = false$ (w functions as an ordinary word):

ii) Draw an ordinary topic: $t \sim Multinomial(\boldsymbol{\theta}_{d(S+1)})$

iii) Draw a word from the ordinary topic t: $w \sim Multinomial(\boldsymbol{\phi}_{t(S+1)})$.

3.2 Decomposed Priors. An important component of the generative process of STDP is to determine the semantic role of a word: whether it functions as sentiment polarity $1, 2, \ldots, S$, or without any polarity (non-sentiment, $S + 1$). The previous LDA-based models (JST, RJST, ASUM) also have this component. They all adopted a flat structure, where the "non-sentiment" ($S+1$, called "neutral" in their models) has an equivalent position of any sentiment polarity. The role of the word is assumed to be simply drawn from an $S + 1$ dimensional discrete distribution of $\tilde{\mathbf{p}} = (p_1, \ldots, p_{S+1})$. In contrast, STDP uses a hierarchical structure where the position of "non-sentiment" is not equivalent to any single polarity but the collection of them. A binary indicator f is drawn first to indicate whether the word functions as a sentiment word or a non-sentiment word. Then another indicator s is drawn to indicate the specific polarity if the word functions as a sentiment word. The $(S + 1)$-dimensional generation distribution $\tilde{\mathbf{p}}$ is decomposed to two levels. The higher one is a 2-dimensional distribution, while the lower one is an S-dimensional distribution.

In the flat structure, to provide the prior for the $S + 1$ dimensional distribution $\tilde{\mathbf{p}}$ is difficult, because the polarity of a word often depends on the topic. The previous models evade the problem and turn to shape the prior of a senti-topic $\boldsymbol{\phi}_{st}$ itself by adding some general polarity words so as to incorporate prior on the discrimination between polarities. The limitation of this approach is analyzed in Section 1 (the second challenge). We overcome the problem by decomposing the flat structure to the two-level hierarchical structure, where we can separately provide prior for the 2-dimensional distribution at the higher layer, i.e., the prior for the binary indicator f. This prior encourages discrimination between sentiment words and non-sentiment words. Compared to the lower layer prior for the S-dimensional distribution to encourage discrimination between specific polarities, this prior for the higher layer is relative easy to obtain.

Table 2: Notations. By default, for a matrix or a vector, we use the corresponding plain letter with subscripts or superscripts to denote a scalar element of it, e.g., θ_{ds}^t is the t-th element of the vector $\boldsymbol{\theta}_{ds}$

D	the number of documents
T	the number of topics
S	the number of sentiments
P	the number of part-of-speech groups
W	the vocabulary size
$\mathcal{S}$	$\{1, 2, \ldots, S\}$, indices of the S sentiment polarities
$\mathcal{S}_+$	$\mathcal{S} \cup \{S + 1\}$, $S + 1$ indicates "non-sentiment"
$\boldsymbol{\theta}_{ds}$	The topic distribution of document d, over the T senti-topics with polarity s
$\boldsymbol{\Theta}$	$[\boldsymbol{\theta}_{ds}]$, the tensor of size $D \times (S + 1) \times T$
$\boldsymbol{\alpha}$	The parameter of the Dirichlet prior for $\boldsymbol{\Theta}$, of size T
$\boldsymbol{\pi}_d$	The sentiment distribution of document d, over the S polarities
$\boldsymbol{\Pi}$	$[\boldsymbol{\pi}_d]$, the matrix of size $D \times S$
$\boldsymbol{\gamma}$	The parameter of the Dirichlet prior for $\boldsymbol{\Pi}$, of size S
$\boldsymbol{\phi}_{st}$	The topic t with sentiment s, a W-dimensional word distribution over the vocabulary
$\boldsymbol{\Phi}$	$[\boldsymbol{\phi}_{st}]$, the tensor of size $(S + 1) \times T \times W$
$\mathbf{b}_s$	The parameter of the Dirichlet prior for $\boldsymbol{\phi}_{st}, \forall t$, of size W
$\mathbf{B}$	$[\mathbf{b}_s]$, the matrix of size $(S + 1) \times W$
$\boldsymbol{\rho}_p$	For part-of-speech group p, the distribution over $\{true, false\}$. ρ_p^{true} indicates the probability of the group generating a sentiment word
$\boldsymbol{\delta}_p$	The parameter of the Beta prior for $\boldsymbol{\rho}_p$, of size 2
$\boldsymbol{\Delta}$	$[\boldsymbol{\delta}_p]$, the matrix of size $P \times 2$
w_i	the i-th word
t_i	the i-th word's topic
s_i	the i-th word's sentiment, taking values from $\mathcal{S}_+$
f_i	the i-th word's flag, taking binary values $true$ or $false$. w_i is a topic-senti word if $f_i = true$, and is not if otherwise
$\mathbf{w}$	the word list for the whole corpus
$\mathbf{t}_{-i}$	the topic assignment for the whole corpus except the i-th word
$\mathbf{s}_{-i}$	the sentiment assignment for the whole corpus except the i-th word
$\mathbf{f}_{-i}$	the flag assignment for the whole corpus except the i-th word
$n_{ds}^{(t)}$	number of words in document d and sentiment $s \in \mathcal{S}_+$ that are in topic t
$n_{ts}^{(w)}$	number of word w in sentiment $s \in \mathcal{S}_+$ that are in topic t
$n_d^{(s)}$	number of words in document d that are in sentiment $s \in \mathcal{S}$
$n_p^{(true)}$	number of words in part-of-speech group p that are topic-senti words
$n_p^{(false)}$	number of words in part-of-speech group p are topic-non-senti words
$n_{\cdot,-i}^{(\cdot)}$	counts without the i-th word

As described in the generation process, we provide the Beta prior $\boldsymbol{\delta}_p$ for each part-of-speech group p. A word whose part-of-speech is in group p uses $\boldsymbol{\delta}_p$ as the prior to generate the Bernoulli distribution $\boldsymbol{\rho}_p$ on the two bins of "sentiment" and "non-sentiment". The Beta prior $\boldsymbol{\delta}_p$ can be set empirically or by analyzing the correlation between part-of-speech tags and the sentiment label on some other corpus with sentiment labels.

There are two benefits for utilizing part-of-speech tags to incorporate priors. First, some previous researchers have observed the close relation between the part-of-speeches and sentiments. Their works to some extend indicates that words with different part-of-speeches are not equally likely to express sentiments[11, 3, 2, 6, 20]. Second, the correlation between the part-of-speeches and sentiments seems not to depend very much on topics or corpus, as it mainly reflects the linguistic behaviors in a language. Thus it is easy to analyze the correlation from some other labeled corpus and directly transfer the correlation to set the prior $\boldsymbol{\delta}_p$.

3.3 Inference Using Gibbs Sampling. We use collapsed Gibbs sampling [7] for inference. For each word w_i, the flag $f_i \in \{true, false\}$ indicates whether w_i functions as a topic-senti word or not, $s_i \in \mathcal{S}$ indicates its polarity if $f_i = true$, and $t_i \in \mathcal{T}$ indicates its topic. The joint posterior of the three variables is as follows.

$$\text{(3.1)} \quad \forall s \in \mathcal{S}, t \in \mathcal{T},$$
$$p(f_i = true, s_i = s, t_i = t | \mathbf{f}_{-i}, \mathbf{s}_{-i}, \mathbf{t}_{-i}, \mathbf{w}) \propto$$
$$\frac{\alpha^{(t)} + n^{(t)}_{ds,-i}}{\sum_{t'}(\alpha^{(t')} + n^{(t')}_{ds}) - 1} \cdot \frac{\beta^{(w)}_s + n^{(w)}_{ts,-i}}{\sum_{w'}(\beta^{(w')}_s + n^{(w')}_{ts}) - 1}$$
$$\cdot \frac{n^{(s)}_{d,-i} + \gamma^{(s)}}{\sum_{s'}(n^{(s')}_d + \gamma^{(s')}) - 1}$$
$$\cdot \frac{n^{(true)}_{p,-i} + \delta^{(true)}_p}{n^{(true)}_p + \delta^{(true)}_p + n^{(false)}_p + \delta^{(false)}_p - 1}$$

and

$$\text{(3.2)} \quad \forall t \in \mathcal{T}, p(f_i = false, t_i = t | \mathbf{f}_{-i}, \mathbf{t}_{-i}, \mathbf{w}) \propto$$
$$\frac{\alpha^{(t)} + n^{(t)}_{d(S+1),-i}}{\sum_{t'}(\alpha^{(t')} + n^{(t')}_{d(S+1)}) - 1} \cdot \frac{\beta^{(w)}_{S+1} + n^{(w)}_{t(S+1),-i}}{\sum_{w'}(\beta^{(w')}_{S+1} + n^{(w')}_{t(S+1)}) - 1}$$
$$\cdot \frac{n^{(false)}_{p,-i} + \delta^{(false)}_p}{n^{(true)}_p + \delta^{(true)}_p + n^{(false)}_p + \delta^{(false)}_p - 1}$$

After burn-in time, the probabilities can be estimated as follows:

(1) The probability of topic t in document d:

$$\text{(3.3)} \qquad \forall s \in \mathcal{S}_+, \theta^{(t)}_{ds} = \frac{\alpha^{(t)} + n^{(t)}_{ds}}{\sum_t(\alpha^{(t)} + n^{(t)}_{ds})}.$$

(2) The probability of word w in topic t and sentiment s:

$$\text{(3.4)} \qquad \forall s \in \mathcal{S}_+, \phi^{(w)}_{ts} = \frac{\beta^{(w)}_s + n^{w}_{ts}}{\sum_w(\beta^{(w)}_s + n^{(w)}_{ts})}.$$

(3) The probability of sentiment s in document d:

$$\text{(3.5)} \qquad \forall s \in \mathcal{S}, \pi^{(s)}_d = \frac{\gamma^{(s)} + n^{(s)}_d}{\sum_s(\gamma^{(s)} + n^{(s)}_d)}.$$

(4) The probability of whether a part-of-speech group p could be used to express sentiment:

$$\text{(3.6)}$$
$$\rho^{(true)}_p \propto n^{(true)}_p + \delta^{(true)}_p, \ \rho^{(false)}_p \propto n^{(false)}_p + \delta^{(false)}_p$$

and $\rho^{(true)}_p + \rho^{(false)}_p = 1$.

3.4 Postprocessing. Some postprocessing is needed to fulfill tasks.

Document-level sentiment label. We use $\boldsymbol{\pi}_d$, the document-level sentiment distribution, to determine the sentiment polarity of document d: we take the polarity $s^\star$ with the highest probability $\pi^{(s^\star)}_d$ as the sentiment label for the review:

$$\text{(3.7)} \qquad s^\star = \arg\max_{s \in \mathcal{S}} \pi^{(s)}_d.$$

Senti-topics and ordinary topics. Senti-topic word distributions are $\{\boldsymbol{\phi}_{st}\}, t \in \mathcal{T}, s \in \mathcal{S}$. Words with high probabilities in a distribution $\boldsymbol{\phi}_{st}$ can be regarded as representative words that describe the senti-topic. The ordinary topic word distributions are $\{\boldsymbol{\phi}_{(S+1)t}\}, t \in \mathcal{T}$. We also take the words with high probabilities in distributions $\boldsymbol{\phi}_{(S+1)t}$ as representative words of the discovered topic. Note that these topics here are different from senti-topics, since they do not carry any sentiments.

Topic-senti words extraction. We need to find the words expressing different sentiment polarities in the same semantic topic. However, in our model the topics are organized by sentiments and there is no one-to-one correspondence guaranteed between the T topics of different sentiments. As a result, we cannot directly tell the set of senti-topics talking about the same semantic topic. The task of topic-senti words extraction requires us to recover the correspondences. We make use of the ordinary topics $\boldsymbol{\phi}_{(S+1)t}$ as bridges to achieve the goal. The idea is straightforward: for each topic t, we take the learnt ordinary topic $\boldsymbol{\phi}_{(S+1)t}$ as the ground truth of the semantic topic, then for each sentiment polarity s, we sweep all the T senti-topics $\{\boldsymbol{\phi}_{st}\}_{t=1}^T$ to find the one with lowest divergences, which will be marked as the corresponding senti-topic (with polarity s) associated with the ordinary topic $\boldsymbol{\phi}_{(S+1)t}$. In this way, we recover the correspondences between different sentiments for all the topics. Then for each topic t and each sentiment polarity s, we take words with high probabilities in the senti-topic $\boldsymbol{\phi}_{st}$, exclude words with high probabilities in the ordinary topic $\boldsymbol{\phi}_{(S+1)t}$ and the words left are considered to be the topic-senti words for topic t with sentiment polarity s.

4 EXPERIMENTAL SETUP

4.1 Data Sets. We do experiments on four data sets. The first two are reviews from the shopping vertical of a commercial search engine, about (1) *Computer and Electronics* referred to as *C&E* and (2) *Local Restaurants and Hotels* referred to as *LH*. The additional two data sets are used in [10], which are publicly available [1]. They are reviews about (3) *Electronic Devices* from Amazon [2] referred to as *A-Ele* and (4) *Restaurants* from Yelp [3] referred to as *Y-Res*. The profiles of the four data sets are reported in Tabel 3.

[1] Available at http://uilab.kaist.ac.kr/research/WSDM11

[2] http://www.amazon.com

[3] http://www.yelp.com

Table 3: Profiles of the data sets

	C&E	LR	A-Ele	Y-Res
Avg.# words/review	14	15	76	153
# reviews	10,000	10,000	24,184	27,458
Vocab. size	7,276	8,532	36,261	39,820

Table 4: Sentiment seed words (partial).

Type (Count)	Seed words
Pos (30)	good, well, nice, excellent, fortunate, amazing
Pos Negation (6)	not_bad, not_disappointed, not_problem
Neg (29)	bad, nasty, poor, unfortunate, annoying, complain
Neg Negation (20)	not_good, not_like, not_well, not_worth

4.2 Preprocessing. We remove punctuation marks, numbers, and all English stop words from the reviews. Porter Stemmer [4] is used for stemming so as to reduce the vocabulary size. Stanford Parser [5] is used to get the part-of-speech tags for all the words in reviews.

In addition, *Negation* is an important preprocessing for sentiment analysis. For example, in a sentence "I do not like the food", "not like" expresses a negative polarity, which is opposite against "like". Without negation preprocessing, it is difficult to figure this out by a *bag-of-words* based model like LDA. To do this, we use Stanford Parser to find the word dependencies in all documents, pick out all "neg(·,·)" dependencies, and add a prefix "not_" to the word in this dependency that is modified by "not".

4.3 Sentiment Priors Construction. We only consider two sentiment polarities, *positive* and *negative*, i.e., $S = 2$. In STDP there are two sources of priors on sentiments.

First is to provide some typical sentiment words to shape the prior $\boldsymbol{B}$ for the senti-topics $\{\boldsymbol{\phi}_{st}\}$ of each polarity $s \in \mathcal{S}$. We choose some general sentiment words which are not dependent on topics. Since we have done negation preprocessing for documents, we could also include some negatived words to our seed word list. A part of seed words and numbers of seed words of each kind are listed in Tabel 4.

Second is to provide prior $\boldsymbol{\Delta}$ on how likely a word functions as a sentiment word. Instead of directly constructing the prior for each word, we construct the prior for each type of part-of-speech. As some part-of-speeches are similar in the tendency of expressing a sentiment, we divide them into five coarse groups, as shown in Tabel 5. Then for each group we empirically set the prior probability of how likely a word from this group functions as a sentiment word.

The specific numerical values for the two prior variables are reported in Section 4.4.

4.4 Models and Parameter Settings. This section explains the parameter settings for all the involved models, JST [13], RJST [14], ASUM [10] and the proposed STDP.

Since we have no prior knowledge on a document's tendency towards different sentiment polarities or topics, we simply use a symmetric $\boldsymbol{\gamma}$ of 0.1 and a symmetric $\boldsymbol{\alpha}$ of $\frac{2.0}{T}$ for all models.

Table 5: Part-of-speech Groups

Group (ID)	Part-of-speeches (POS) in the group
Adjective-related (1)	JJ, JJS, JJR
Adverb-related (2)	RB, RBS, RBR
Verb-related (3)	VB, VBD, VBG, VBN, VBP, VBZ
Noun-related (4)	NN, NNP, NNS, NNPS
Others (5)	POS's not included in the previous 4 groups

The parameter $\boldsymbol{B}$ indicates how likely words are positive or negative. This parameter is where we incorporate our prior knowledge of sentiment seed words (Section 4.3). In JST, RJST and ASUM, $\boldsymbol{B}$'s have the same size of $S \times W$, we set their $\boldsymbol{B}$'s to be the same: For positive sentiment s, we set elements in $\boldsymbol{\beta}_s$ to be 0.1 for positive sentiment seed words, 0.001 for negative sentiment seed words and 0.01 for other words; Similarly, for negative sentiment s we set elements of $\boldsymbol{\beta}_s$ to be 0.1 for negative sentiment seed words, 0.001 for positive sentiment seed words and 0.01 for other words. In STDP, the dimension of $\boldsymbol{B}$ is $(S + 1) \times W$, suggesting that besides the positive and negative senti-topics, we have an additional word distribution on ordinary topic. We set the first S rows of $\boldsymbol{B}$ the same as JST, RJST and ASUM, and for the ordinary topic, the $(S + 1)$-th row, we setelements of it to be 0.001 for all sentiment seed words, and 0.01 for other words.

The other parameter for STDP is $\boldsymbol{\Delta}$ of size $P \times 2$, where $P = 5$ since we group all part-of-speeches into 5 groups. For each part-of-speech group p (the index variable p corresponds to the ID of the group, see Tabel 5), $\boldsymbol{\delta}_p$ is the parameter for a Beta distribution, a 2-dimensional vector. We represent

$$\boldsymbol{\delta}_p = |\boldsymbol{\delta}_p|(\sigma_p, 1 - \sigma_p), \tag{4.8}$$

where $|\boldsymbol{\delta}_p|$ represents the strength of the prior, $0 < \sigma_p < 1$ indicates the shape of the prior, i.e., how likely the group p generates a sentiment word. For all groups, we set the same strength $|\boldsymbol{\delta_p}| = 10$. And we set the shape parameters $\boldsymbol{\sigma} = (\sigma_1, \ldots, \sigma_5) = (0.9, 0.9, 0.7, 0.9, 0.1)$. Section 5.1.3 will show the impact of different settings of this prior.

5 RESULTS & ANALYSIS

We report and analyze the experimental results from four aspects: document level sentiment analysis, senti-topic discovery, topic discovery and topic-senti word extraction.

5.1 Document Level Sentiment Analysis. All the four models are able to predict the sentiment label of a review document. For all the four data sets we used, we have a document level binary sentiment label (*positive* or *negative*) for each review. Therefore we use the label to measure the prediction accuracy. This document level sentiment accuracy is one of the most important criterion for a sentiment model as it quantitatively and clearly measures the ability of the model to discriminate between different sentiment polarities.

Note that for RJST, we need some postprocessing to get the document level sentiment label. For each document d,

[4] http://tartarus.org/ martin/PorterStemmer/

[5] http://nlp.stanford.edu/software/lex-parser.shtml

Table 6: Document level sentiment prediction accuracy (%) on C&E and LH. $T = 30$.

	JST	RJST	ASUM	STDP
C&E	74.41	58.27	68.59	**77.35**
LH	78.22	54.31	72.59	**82.40**

Table 7: Document level sentiment prediction accuracy (%) on A-Ele and Y-Res. $T = 30$.

	LP-Uni	LP-Bi	ASUM	JST(Redo)	STDP
A-Ele	71	79	84	78	**86**
Y-Res	81	87	86	80	**88**

RJST outputs a topic distribution $\boldsymbol{\theta}_d$ on the T topics, and a sentiment distribution $\boldsymbol{\pi}_{dt}$ on the S polarities for each topic t. Then the document level sentiment distribution $\boldsymbol{\pi}_d$ is obtained by straightforward probability marginalization:

$$\forall s \in \mathcal{S},\ \pi_d^{(s)} = \sum_t \theta_d^{(t)} \pi_{dt}^{(s)}. \tag{5.9}$$

Having $\boldsymbol{\pi}_d$, the polarity label is obtained by Equation 3.7.

5.1.1 Results with different models. Results on the two data sets *C&E* and *LH* are reported in Tabel 6. STDP outperforms all the other three models. The performance of RJST is the worst with no surprise, due to reasons shown in Section 2. This drawback is conquered in STDP (also in JST and ASUM) by directly modeling sentiment distributions over each document. We also see that ASUM is outperformed by JST, which seems to contradict the claims in [10] that ASUM is better than JST. The reason is that, ASUM assumes "words from one sentence are picked from the same senti-topic". The assumption is reasonable for long reviews with many sentences, where one sentence could be regarded as the unit of senti-topic. However, in *C&E* and *LH*, reviews are very short, most of which consist of one or two sentences. In this case using a sentence as a unit causes high bias hence the assumption is too strong. Thus as expected, both JST and STDP outperform ASUM because they do not make such assumptions. Last we see that STDP outperforms JST by approximately 3% and 4% on C&E and LH respectively, revealing the benefits of the decomposed priors.

The two data sets *A-Ele* and *Y-Res* contain long reviews. Hence the results on them help us further investigate whether the proposed STDP model is vulnerable and bias to short reviews. The results on *A-Ele* and *Y-Res* are reported in Tabel 7. As *A-Ele* and *Y-Res* are the data sets that ASUM experiments on [10], in Tabel 7 the results of ASUM are directly copied from the original paper of ASUM [10]. In addition, in the paper the author found another supervised method LinePipe [1] competes ASUM, we also copy the results of the method to the table (methods LP-Uni and LP-Bi) for reference, where LP-Uni and LP-Bi stand for the unigram and bigram versions of LinePipe. We need to point out that the result of JST is from our re-implementation as we found our implementation produces much better results than that reported in the ASUM paper [10]. For the proposed method STDP, for a fair comparison, we use the same set of sentiment seed words and the same number of topics ($T = 30$) as those used in ASUM paper. The results show that, JST performs slightly better than it does on short reviews in *C&E* and *LH*, but is outperformed by ASUM by nearly 6%. On the other hand, STDP outperforms all the other models. Most surprisingly, it even slightly outperforms LingPipe's (*LP-Uni* and *LP-Bi*), which are fully supervised. This experiment shows that, STDP is able to perform well on both long reviews and short reviews, which is not achieved by either ASUM or JST.

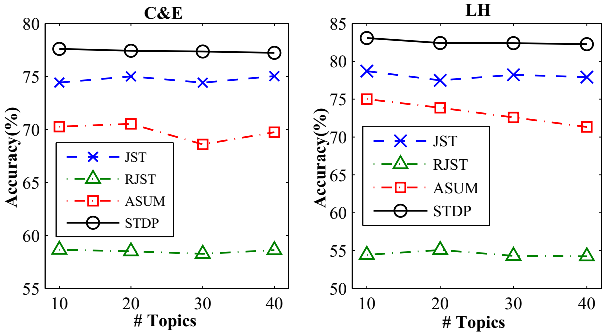

Figure 2: The impact of the number of topics on the sentiment prediction accuracy on data sets *C&E* and *LH*.

5.1.2 Impact of the number of topics T. To investigate the impact of the number of topics on different models, on data sets *C&E* and *LH*, we sweep the number of topics T in $\{10, 20, 30, 40\}$, and plot the results in Figure 2. STDP still outperforms all the other three models in all the settings of number of topics. Among the models, ASUM is most sensitive to the number of topics, especially on data set *LH*. In contrast, STDP, JST, and RJST are not sensitive to this parameter.

5.1.3 Impact of the sentiment prior As explained in Section 4.4, $\boldsymbol{\Delta}$ controls a part-of-group's prior likelihood to generate a sentiment word. For each group p, we decouple the parameter $\boldsymbol{\delta}_p$ to the strength parameter $|\boldsymbol{\delta}_p|$ and the shape parameter σ_p by the representation $\boldsymbol{\delta}_p = |\boldsymbol{\delta}_p| \cdot (\sigma_p, 1 - \sigma_p)$ of Equation 4.8. In previous experiments, we empirically set $|\boldsymbol{\delta}_p| = 10, \forall p$, and $\boldsymbol{\sigma} = (\sigma_1, \ldots, \sigma_5) = (0.9, 0.9, 0.7, 0.9, 0.1)$. To investigate the impact of this prior, on data set *LH*, with fixed strength $|\boldsymbol{\delta}_p| = 10$, we sweep the shape parameter σ_p in $\{0.1, 0.3, 0.5, 0.9\}$, and report the results in Tabel 8.

By increasing σ_5 (for the group *Other*), the accuracy of STDP continues decreasing from 83.03% to 82.67%, indicating that encouraging words in such part-of-speech group as sentiment words has a negative impact on the performance. By changing the value of σ_4 (for the group *Noun-related*), the accuracy slightly decreases and then fluctuates, showing that *Noun-related* words have a little impact on the model. When setting $\sigma_3 = 0.7$ (for the group *Verb-related*) the performance goes to 83.08%, which is the best among all the settings, but then it drops to 82.58% for $\sigma_3 = 0.1$. This indicates that *Verb-related* words have positive impact in sentiment analysis. By decreasing σ_1 (for the group *Adj-*

Table 8: Performance of STDP on data set LH under different settings of σ.

	σ_1	σ_2	σ_3	σ_4	σ_5	Accuracy(%)
Impact of σ_5	0.9	0.9	0.9	0.9	0.9	82.67
	0.9	0.9	0.9	0.9	0.7	82.65
	0.9	0.9	0.9	0.9	0.5	82.88
	0.9	0.9	0.9	0.9	0.3	82.91
	0.9	0.9	0.9	0.9	0.1	83.03
Impact of σ_4	0.9	0.9	0.9	0.7	0.1	82.91
	0.9	0.9	0.9	0.5	0.1	82.84
	0.9	0.9	0.9	0.3	0.1	82.88
	0.9	0.9	0.9	0.1	0.1	82.84
Impact of σ_3	**0.9**	**0.9**	**0.7**	**0.9**	**0.1**	**83.08**
	0.9	0.9	0.5	0.9	0.1	82.77
	0.9	0.9	0.3	0.9	0.1	82.65
	0.9	0.9	0.1	0.9	0.1	82.58
Impact of σ_2	0.9	0.7	0.7	0.9	0.1	82.92
	0.9	0.5	0.7	0.9	0.1	82.86
	0.9	0.3	0.7	0.9	0.1	82.79
	0.9	0.1	0.7	0.9	0.1	82.60
Impact of σ_1	0.7	0.9	0.7	0.9	0.1	82.82
	0.5	0.9	0.7	0.9	0.1	82.74
	0.3	0.9	0.7	0.9	0.1	82.55
	0.1	**0.9**	**0.7**	**0.9**	**0.1**	**82.23**

related) and σ_2 (for the group *Adv-related*), the accuracy keeps dropping until it reaches the worst of 82.23%, showing that most *Adj-Related* and *Adv-Related* words have positive impact on STDP.

However, from the absolute values, we can see the performance is robust to the parameter. This is because we use the prior in a full Bayesian approach, where $\boldsymbol{\delta}_p$ just plays as the prior for the distribution $\boldsymbol{\rho}_p$, which is the real one directly determining the probability of generating a sentiment word from group p.

5.2 Senti-Topic Discovery. We show the senti-topics discovered by STDP by listing the representative words. We exclude sentiment seed words from our lists, as we consider they are general instead of topic specific. Limited by space, we only list a part of topics and words on data sets *C&E* and *LH*, as shown in Tabel 9. After post checking the words in a sent-topic, we manually add a label for each senti-topic. The sentiment polarity of a senti-topic may be positive or negative, shorted as *pos* and *neg* respectively.

In word lists shown, we not only find topic-senti words like *slow*, *fast*, *easy*, etc, but also some topic-non-senti words. For example, We find word *usb* under senti-topic *hard/software(neg)*, and *baker* under senti-topic *equipment(neg)*. These words do not express sentiments without context, but have a strong indication of what the topic is about. The reason that they appear in senti-topics is that they frequently co-occur with those topic-senti words, indicating that the sentiments expressed by topic-senti words are essentially towards these topic-non-senti words. We could say that most reviews have negative sentiments over *usb* when talking about *hard/software* and *baker* when talking about the *equipment* of a restaurant.

Interestingly, we find some pairs of antonyms under the same senti-topic, as emphasized in red in Tabel 9. Intuitively we may feel antonyms are used to express opposite sentiments, but this seems not to be true. As indicated in the table, people may use "*too hot*" or "*too cold*" to complain about room temperature, and "*modern*" or "*antique*" to praise the room decoration. Being neutral under some topics may meet with great favor, while under other topics may not.

5.3 Topic Discovery & Topic-Senti Words Extraction. We list words with high probabilities in $\phi_{(S+1)t}$ that are recognized as ordinary topic words, and list topic-senti words extracted using the way described in the last paragraph of Section 3.4. The results on data sets *C&E* and *LH* are shown in Tabel 10. The number of topics is set to 20.

We pick 2 sets of words, each containing topic-non-senti words describing what the topic is about (i.e., words extracted in topic discovery), and positive and negative words describing how sentiments over topics are expressed (topic-senti words extracted).

The first set contains words on the topic *printers and pictures* as can be drawn from the first column; in the second column we can see that people use topic-senti words like *clear, bright, sharp, fast*, etc. to express positive sentiment over the quality of the pictures and speed of printing; in the third column we see that words like *black, slow, white* are extracted, indicating that these words are topic-senti words used to express negative sentiment over the topic.

The second set of words shows how topic of *battery* is described. As we can see, positive topic-senti words like *small, easily, lightweight* and negative topic-senti words like *fast, noisy, low, weak* are extracted.

Nearly all the topic words extracted are topic words that are not biased to sentiment words, showing the pureness of discovered topics.

One interesting thing is also shown in these two sets of words: word *fast* (emphasized in red in both sets) is extracted in the positive word list under topic *printer and pictures*, but is recognized as negative topic-senti word under topic *battery*, suggesting people may use *fast* to express different sentiments under different topics. For example, people may use *fast* to praise the speed of printing, and to criticize that the power of battery drops fast. This clearly shows that, STDP is able to extract words that expresses different sentiment under different topics, i.e., those topic-senti words, and perform well in word-level sentiment analysis.

6 CONCLUSION

In this paper, we propose an unsupervised generative model (STDP) for jointly mining topics, sentiments, and their associations from online reviews. In the model we decompose the sentiment generative process into a two-level hierarchy, and provide priors to encourage the discrimination between sentiment words and non-sentiment words. We construct this prior according to the part-of-speech tag of a word. The experiments on real review data sets show that STDP is effective in discovering topics & senti-topics and extracting topic-senti words. Most importantly, STDP keeps good discriminative performance on sentiment analysis: on docu-

Table 9: Senti-Topics discovered by STDP on data set *C&E* and *LH* . Topic names for each set of words are manually labeled. The number of topics is set to 20.

Data set	Topic(polarity)	Representative words
C&E	perform(neg)	slow, low, seem, weak, little, bit, flimsy, feel, run, maybe, real, somewhat, due, slower, optical
	perform(pos)	much, better, seem, expect, faster, compare, easier, wish, regular, lack, sound, improved, previous, take, little
	cellphone(neg)	hard, sometime, change, remote, lock, make, actual, press, take, push, somewhat, tell, design, poorly, universal
	cellphone(pos)	easy, small, compact, button, carry, easily, hold, fit, clear, lightweight, sharp, around, actual, scroll, type
	(h/s)ware(neg)	usb, cause, crash, wireless, not_had, either, previous, found, entire, attach, not_come, properly, buggy, set, onboard
	(h/s)ware(pos)	fast, reliable, look, extremely, feel, solid, quiet, large, stable, durable, plastic, strong, black, especially, huge
	camera(neg)	complete, make, useless, scan, cause, include, print, loud, within, run, run, across, frequent, come, wait, download
	printer(pos)	take, able, hard, find, took, hook, digital, video, allow, learn, figure, turn, transfer, picture, plus
LH	room(neg)	dirty, smell, small, look, cover, stain, musty, gross, dark, dingy, peel, smelly, outdated, felt, filthy,...
	room(neg)	little, felt, hot, dirty, cold, whole, last, fall, shabby, agree, gross, end, slept, begin, yet,...
	equipment(neg)	broken, felt, little, look, fall, shabby, third, worn, pull, over, rent, behind, end, whole, tired,...
	service(neg)	order, came, took, slow, real, serve, quit, arrive, final, extreme, huge, sat, leave, empty, ignore,...
	room(pos)	decoration, private, feature, original, antique, furnish, modern, victorian, boast, beautiful, Italian, local, restore,...
	food(pos)	fresh, serve, sit, big, try, delicious, overlook, follow, often, fill, exquisite, full, stun, baker, hearty,...
	location(pos)	walk, location, within, close, right, across, easy, central, short, nearby, history, near, downtown, easily, major,...
	service(pos)	wonderful, feel, welcome, special, make, luxury, found, expect, grand, superb, comfy, cozy, treat, celebrate, everywhere,...

Table 10: Topic-Senti words discovered by STDP on data set *C&E* and *LH*. Words in brackets word lists are topic-senti words that should be excluded. The number of topics is set to 20.

Data set	Polarity	Representative words
C&E	Topic-non-senti	printer, print, photo, quality, paper, ink, color, hp, page, document, epson, scan, cost, machine, canon
	Positive	clear, sound, incredible, (print), cirspy, bright, sharp, fast, clean, superb, (product), super, crystal, include, absolute
	Negative	include, (print), within, black, total, found, want, run, instead, slow, (scan), white, sell, constantly, (color)
LH	Topic-non-senti	battery, life, case, ipod, thing, charger, hour, cover, door, plastic, something, charge, pocket, place, cord
	Positive	fit, small, easily, really, hold, look, (plastic), lightweight, compact, adjust, quit, thin, flat, wide, finish
	Negative	little, get, (bit), (somewhat), noisy, (sometime), flimsy, low, weak, found, limit, loose, come, fast, confuse

ment level sentiment prediction accuracy, STDP outperforms all other compared methods, not only on short reviews but also on long reviews, which cannot be achieved by any previous methods.

References

[1] Alias-i. lingpipe 4.0.1. http://alias-i.com/lingpipe, 2008.

[2] F. Benamara, C. Cesarano, A. Picariello, D. Reforgiato, and V. Subrahmanian. Sentiment analysis: Adjectives and adverbs are better than adjectives alone. In *ICWSM*, 2007.

[3] S. Bethard, H. Yu, A. Thornton, V. Hatzivassiloglou, and D. Jurafsky. Automatic extraction of opinion propositions and their holders. In *AAAI*, 2004.

[4] D. Blei, A. Ng, and M. Jordan. Latent dirichlet allocation. *J. Mach. Learn. Res.*, 3:993–1022, 2003.

[5] S. Brody and N. Elhadad. An unsupervised aspect-sentiment model for online reviews. In *HLT-NAACL*, pages 804–812, 2010.

[6] P. Chesley, B. Vincent, L. Xu, and R. Srihari. Using verbs and adjectives to automatically classify blog sentiment. In *AAAI-CAAW*, 2006.

[7] T. Griffiths and M. Steyvers. Finding scientific topics. *PNAS*, 101(Suppl 1):5228–5235, 2004.

[8] T. Griffiths, M. Steyvers, D. Blei, and J. Tenenbaum. Integrating topics and syntax. In *NIPS*, volume 17, pages 537–544, 2005.

[9] T. Hofmann. Probabilistic latent semantic indexing. In *SIGIR*, pages 50–57. ACM, 1999.

[10] Y. Jo and A. Oh. Aspect and sentiment unification model for online review analysis. In *WSDM*, pages 815–824. ACM, 2011.

[11] J. Klavans and M. Kan. Role of verbs in document analysis. In *ACL*, pages 680–686, 1998.

[12] H. Lakkaraju, C. Bhattacharyya, I. Bhattacharya, and S. Merugu. exploiting coherence for the simultaneous discovery of latent facets and associated sentiments. In *SDM*, 2011.

[13] C. Lin and Y. He. Joint sentiment/topic model for sentiment analysis. In *CIKM*, pages 375–384. ACM, 2009.

[14] C. Lin, Y. He, R. Everson, and S. Ruger. Weakly supervised joint sentiment-topic detection from text. *IEEE T. Knowl. Data. En.*, 24(6):1134–1145, 2012.

[15] Q. Mei, X. Ling, M. Wondra, H. Su, and C. Zhai. Topic sentiment mixture: modeling facets and opinions in weblogs. In *WWW*, pages 171–180. ACM, 2007.

[16] S. Moghaddam and M. Ester. Aspect-based opinion mining from online reviews. *Tutorial at SIGIR*, 2012.

[17] B. OConnor, R. Balasubramanyan, B. Routledge, and N. Smith. From tweets to polls: Linking text sentiment to public opinion time series. In *ICWSM*, pages 122–129, 2010.

[18] B. Pang, L. Lee, and S. Vaithyanathan. Thumbs up?: sentiment classification using machine learning techniques. In *EMNLP*, pages 79–86, 2002.

[19] I. Titov and R. McDonald. A joint model of text and aspect ratings for sentiment summarization. In *ACL*, pages 308–316, 2008.

[20] J. Yi, T. Nasukawa, R. Bunescu, and W. Niblack. Sentiment analyzer: Extracting sentiments about a given topic using natural language processing techniques. In *ICDM*, pages 427–434. IEEE, 2003.

[21] J. Zhu and E. Xing. Conditional topic random fields. In *ICML*, 2010.