Aspect and Sentiment Unification Model
for Online Review Analysis

Yohan Jo
Department of Computer Science
KAIST
Daejeon, Korea
yohan.jo@kaist.ac.kr

Alice Oh
Department of Computer Science
KAIST
Daejeon, Korea
alice.oh@kaist.edu

ABSTRACT

User-generated reviews on the Web contain sentiments about detailed aspects of products and services. However, most of the reviews are plain text and thus require much effort to obtain information about relevant details. In this paper, we tackle the problem of automatically discovering what aspects are evaluated in reviews and how sentiments for different aspects are expressed. We first propose *Sentence-LDA* (SLDA), a probabilistic generative model that assumes all words in a single sentence are generated from one aspect. We then extend SLDA to *Aspect and Sentiment Unification Model* (ASUM), which incorporates aspect and sentiment together to model sentiments toward different aspects. ASUM discovers pairs of {aspect, sentiment} which we call *senti-aspects*. We applied SLDA and ASUM to reviews of electronic devices and restaurants. The results show that the aspects discovered by SLDA match evaluative details of the reviews, and the senti-aspects found by ASUM capture important aspects that are closely coupled with a sentiment. The results of sentiment classification show that ASUM outperforms other generative models and comes close to supervised classification methods. One important advantage of ASUM is that it does not require any sentiment labels of the reviews, which are often expensive to obtain.

Categories and Subject Descriptors

H.3.3 [**Information Storage and Retrieval**]: Information Search and Retrieval—*Retrieval models*; G.3 [**Probability and Statistics**]: *Probabilistic algorithms*; I.2.7 [**Artificial Intelligence**]: Natural Language Processing—*Text analysis*

General Terms

Algorithms, Experimentation

Keywords

aspect detection, sentiment analysis, topic modeling

Amazon Verified Purchase (What's this?)
This review is from: HP Pavilion DV4-2161NR 14.1-Inch Laptop (Digital Plaid) (Personal Computers)
I've owned my computer for almost a week now, and I'm absolutely loving it. For the money I almost bought an Hp with a T6600 processor and 320GB hard drive etc etc. I bought this for the same price as the other hp and got MORE memory and a better processor. Why wouldn't you want an upgrade of 180 GB hard drive for free? If your a student and like to have lots of music, videos and play light gaming and still have tons of room for your homework and projects, this is a great computer.

I also love how its 14" monitor, so I can take it anywhere and fit it into any bag. It has a very sleek and glossy exterior. A little bit of work to keep clean, but that's no reason not to buy a computer. The speakers are great for a laptop. I was actually surprised how clear sounds and music

Figure 1: Example laptop review from Amazon.com

1. INTRODUCTION

The Web has an overwhelming amount of reviews of products, restaurants, books, and many other types of tangibles and intangibles. In those reviews, people praise and criticize a variety of *aspects* of the target of the review, such as the waiting time of a restaurant or the noise level of a vacuum cleaner. In Figure 1, the reviewer evaluates aspects of a laptop such as the price, monitor size, and sound. Although some Websites (e.g., TripAdvisor) are specifically designed for user reviews with a predefined evaluation form, most users express their opinion in online communities and personal blogs using plain text without any structure.

One big problem is to find aspects that users evaluate in reviews. From the perspective of a user reading the reviews to get information about a product, the evaluations of the specific aspects are just as important as the overall rating of the product. A user looking to buy a digital camera may want to know what a review says about the photo quality, brightness of lens, and shutter speed of a Panasonic Lumix, not just whether the review recommends the camera. Although sometimes the aspect information is available, it is unlikely to be a comprehensive set of all aspects that are evaluated in the reviews. Another important task in review analysis is discovering how opinions and sentiments for different aspects are expressed. The cell phone's battery lasts "long", a laptop's screen "reflects", and a restaurant's server is "attentive". These are sentiment words at the level of the aspect. Previous efforts have mostly focused on sentiment words at the level of the domain (e.g., electronics, movies, restaurants).

We tackle these two problems at once with a unified generative model of aspect and sentiment. Probabilistic topic models are suitable for the following two reasons: first, they provide an unsupervised way of discovering topics from documents (or aspects from reviews), and second, they result in language models that explain how much a word is related to each topic and possibly to a sentiment. We take the latent Dirichlet allocation (LDA) model [4] and adapt it to match the granularity of the discovered topics to the details of the reviews. In addition, we incorporate sentiment into our unified model so that the resulting language models represent the probability distributions over words for various pairs of aspect and sentiment.

An important observation is that in reviews, one sentence tends to represent one aspect and one sentiment. Figure 1 shows an example that supports this. The review is evaluating several aspects including the price, free upgrade, size, and sound, and each sentence expresses sentiment about one aspect. In the first sentence in the second paragraph, the words "monitor" and "bag" co-occur. In general, these two words are not closely related, but the co-occurrence of them signals that this sentence is evaluating the size of the monitor. We use this observation in our models.

In this paper, we propose two models: *Sentence-LDA* (SLDA) and *Aspect and Sentiment Unification Model* (ASUM). SLDA and ASUM model the generative process of reviews. Based on the observation above, SLDA and ASUM constrain that all words in a single sentence be generated from one topic. ASUM is an extension of SLDA into which sentiment is incorporated. In ASUM, the words in a sentence are generated from the same pair of aspect and sentiment, which we call *senti-aspect*.

We applied our models to various tasks, and the experiments show that our models perform well in the following.

- **Aspect discovery**: SLDA finds aspects that match the details of the reviews better than LDA.

- **Senti-aspect discovery**: ASUM finds senti-aspects that reflect both aspect and sentiment, and some aspects strongly related to a sentiment are discovered only by ASUM, not by SLDA.

- **Aspect-specific sentiment words**: ASUM takes a set of general affective and evaluative words and finds aspect-specific evaluative words. This is simple sentiment word expansion and adaptation without labeled data.

- **Sentiment classification**: Although ASUM is not specifically designed for sentiment classification, ASUM performs better than other generative models and almost matches the best performance of a supervised classifier. Much of Web review data is unlabeled, so unsupervised classification of sentiment is an important problem.

2. TERMINOLOGY

This section defines the terminology used in this paper.

- **topic**: a multinomial distribution over words that represents a coherent concept in text.

- **aspect**: a multinomial distribution over words that represents a more specific topic in reviews, for example, "lens" in camera reviews.

- **senti-aspect**: a multinomial distribution over words that represents a pair of aspect and sentiment, for example, "screen, positive" in a laptop review.

- **affective word**: a word that expresses a feeling, for example "satisfied", "disappointed".

- **evaluative word**: a word that expresses sentiment by evaluating an aspect, for example, "excellent", "nice".

- **general evaluative word**: an evaluative word that expresses a consistent sentiment every time it is used, for example, "good", "bad".

- **aspect-specific evaluative word**: an evaluative word that may express different sentiments depending on the aspect, for example, a "small" font size on a monitor that is hard to read vs. a "small" vacuum that is portable.

- **sentiment word**: a word that conveys sentiment. It is either an affective word, general evaluative word, or aspect-specific evaluative word.

3. RELATED WORK

In this section, we describe related research fields of *aspect discovery* and *domain adaptation of sentiment words*. We also discuss several *unified models of topic and sentiment* and compare them with ASUM in details.

A widely used approach in aspect discovery is to extract a set of frequently occurring noun phrases (NP) as aspect candidates and then retain only relevant ones by applying various filtering methods [3, 12, 19]. The NP detection is a complex process that may be error-prone and pose difficulties for cross-domain and cross-lingual applications. Another approach employs topic modeling, for example, fitting the latent Dirichlet allocation (LDA) model [4] to sentences instead of documents [6, 26]. This approach does not consider the relationships among sentences, thus ignoring the fact that the same aspect may have quite different word usages in different sentences. Another way of using a topic model distinguishes between broad topics and fine-grained ratable topics [23]. Our models do not discriminate the two types of topics, but have a simpler and more intuitive generative process to discover evaluative aspects in reviews.

Research on domain adaptation of sentiment words can be categorized into domain-to-domain adaptation and general-to-domain adaptation. The domain-to-domain adaptation aims to obtain sentiment words in one domain by utilizing a set of known sentiment words in another domain [2, 5, 18]. The general-to-domain adaptation takes a set of known general sentiment words and learns domain-specific sentiment words [7, 13]. For sentiment seed words, existing sentiment word lexicons (e.g., SentiWordNet) can be used or a new set of words may be obtained by using sentiment propagation techniques [14, 17, 20]. ASUM starts from a small set of general sentiment words and finds sentiment words related to specific aspects.

Several unified models of topic and sentiment have been proposed [15, 16, 22]. They extend basic topic models that do not consider sentiment [4, 11] to explain the generative process of opinionated documents such as reviews and blogs. All these topic models posit that a document is a mixture over an underlying set of topics, and, in turn, a topic is represented as a multinomial distribution over words.

Topic Sentiment Mixture (TSM) model [16] represents sentiment as a language model separate from topics, and each word comes from either topics or sentiment. This separation cannot explain the intimate interplay between a topic and a sentiment. For ASUM, in contrast, a pair of topic and sentiment is represented as a single language model, where a word is more probable as it is closely related to both the topic and the sentiment. This provides a sound explanation of how much a word is related to certain topic and sentiment.

Multi-Aspect Sentiment (MAS) model [22] differs from the other models in that it focuses on modeling topics to match a set of predefined aspects that are explicitly rated by users in reviews. Sentiment is modeled as a probability distribution over different sentiments for each of the aspects, and this distribution is derived from a weighted combination of discovered topics and words. To fit the discovered topics and sentiment to the predefined aspects and their ratings, MAS requires training data that are rated by users for each aspect. ASUM does not use any user-rated training data, which are often expensive to obtain.

Joint Sentiment/Topic (JST) model [15] takes the most similar approach to ours. Sentiment is integrated with a topic in a single language model. JST does not limit individual words JST is different from ASUM in that individual words may come from different language models. In contrast, ASUM constrains the words in a single sentence to come from the same language model, so that each of the inferred language models is more focused on the regional co-occurrences of the words in a document. Both JST and ASUM make use of a small seed set of sentiment words, but the exploitation is not explicitly modeled in JST. ASUM integrates the seed words into the generative process, and this provides ASUM with a more stable statistical foundation.

It is difficult to compare the aspects and sentiments found by the different models. We carried out sentiment classification for a quantitative comparison, although it is not the main goal of the models. The results, presented in Section 6.4, show that ASUM outperforms TSM and JST.

4. MODELS

We propose two generative models that extend LDA, one of the most widely used probabilistic topic models [4]. Our goal is to discover topics that match the aspects discussed in reviews.

4.1 Sentence-LDA

In LDA, the positions of individual words are neglected for topic inference. As discussed in previous work [25], this property may not always be appropriate. In reviews, words about an aspect tend to co-occur within close proximity to one another. SLDA imposes a constraint that all words in a sentence are generated from one topic. This is not always true, but the constraint holds up well in practice. The graphical representation of SLDA is shown in Figure 2(a) and the notations are explained in Table 1.

In SLDA, the generative process is as follows:

1. For every aspect z, draw a word distribution $\phi_z \sim$ Dirichlet(β)

2. For each review d,

    (a) Draw the review's aspect distribution $\theta_d \sim$ Dirichlet(α)

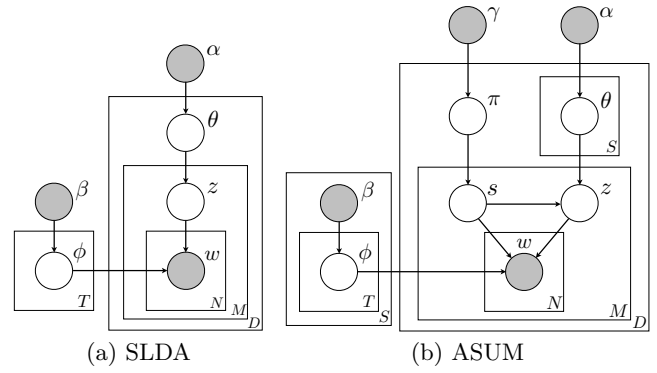   (b) For each sentence,

(a) SLDA   (b) ASUM

Figure 2: Graphical representation of SLDA and ASUM. Nodes are random variables, edges are dependencies, and plates are replications. Only shaded nodes are observable.

    i. Choose an aspect $z \sim$ Multinomial(θ_d)

    ii. Generate words $w \sim$ Multinomial(ϕ_z)

We use Gibbs sampling [10] to estimate the latent variables θ and ϕ. At each transition step of the Markov chain, the aspect of the ith sentence, z_i, is drawn from the conditional probability

$$\mathrm{P}\left(z_i = k | \mathbf{z}_{-i}, \mathbf{w}\right) \propto \frac{C_{dk}^{DT} + \alpha_k}{\sum_{k'=1}^{T} C_{dk'}^{DT} + \alpha_{k'}}$$

$$\frac{\Gamma(\sum_{w=1}^{W} C_{kw}^{TW} + \beta_w)}{\Gamma(\sum_{w=1}^{W} (C_{kw}^{TW} + \beta_w) + m_i)} \prod_{w=1}^{W} \frac{\Gamma(C_{kw}^{TW} + \beta_w + m_{iw})}{\Gamma(C_{kw}^{TW} + \beta_w)}.$$

The notations are described in Table 1, with a minor exceptional use of notation that C_{dk}^{DT} and C_{kw}^{TW} in this expression exclude sentence i.

The approximate probability of aspect k in review d is

$$\theta_{dk} = \frac{C_{dk}^{DT} + \alpha_k}{\sum_{k'=1}^{T} C_{dk'}^{DT} + \alpha_{k'}}.$$

The approximate probability of word w in aspect k is

$$\phi_{kw} = \frac{C_{kw}^{TW} + \beta_w}{\sum_{w'=1}^{V} C_{kw'}^{TW} + \beta_{w'}}.$$

4.2 Aspect and Sentiment Unification Model

ASUM is an extension of SLDA that incorporates both aspect and sentiment. ASUM models the generative process of a review as illustrated in the following scenario of writing a review. A reviewer first decides to write a review of a restaurant that expresses a distribution of sentiments, for example, 70% satisfied and 30% unsatisfied. And he decides the distribution of the aspects for each sentiment, say 50% about the service, 25% about the food quality, and 25% about the price for the positive sentiment. Then he decides, for each sentence, a sentiment to express and an aspect for which he feels that sentiment. For example, he writes that he is satisfied with the friendly service of the restaurant. The graphical representation of ASUM is shown in Figure 2(b). Formally, the generative process is as follows:

1. For every pair of sentiment s and aspect z, draw a word distribution $\phi_{sz} \sim$ Dirichlet(β_s)

Table 1: Meanings of the notations

D	the number of reviews
M	the number of sentences
N	the number of words
T	the number of aspects
S	the number of sentiments
V	the vocabulary size
w	word
z	aspect
s	sentiment
ϕ	multinomial distribution over words
θ	multinomial distribution over aspects
π	multinomial distribution over sentiments
$\alpha_{(k)}$	Dirichlet prior vector for θ
$\beta_{(w)}, \beta_{j(w)}$	Dirichlet prior vector for ϕ (for sentiment j)
$\gamma_{(j)}$	Dirichlet prior vector for π
z_i	the aspect of sentence i
s_i	the sentiment of sentence i
$\mathbf{z}_{-i}$	the aspect assignments for all sentences except sentence i
$\mathbf{s}_{-i}$	the sentiment assignments for all sentences except sentence i
$\mathbf{w}_i$	the word list representation of sentence i
$\mathbf{w}$	the word list representation of the corpus
C_{dk}^{DT}	the number of sentences that are assigned aspect k in review d
C_{kw}^{TW}	the number of words that are assigned aspect k
C_{dj}^{DS}	the number of sentences that are assigned sentiment j in review d
C_{djk}^{DST}	the number of sentences that are assigned sentiment j and aspect k in review d
C_{jkw}^{STW}	the number of words that are assigned sentiment j and aspect k
$m_{i(w)}$	the number of total words (or word w) in sentence i

Table 2: The properties of the data sets

	ELECTRONICS	RESTAURANTS
# of reviews	24,184	27,458
# of reviews with 4+ stars	72%	68%
Avg. # of words/review	76	153
Avg. # of sentences/review	12	12

be more probable in the corresponding senti-aspects. Symmetric β, which was often used in previous work, does not utilize this prior knowledge. A similar unified model [15] incorporated sentiment information at the initialization step of Gibbs sampling, but the effect becomes weak as the sampling progresses.

The latent variables θ, π, and ϕ are inferred by Gibbs sampling as in Section 4.1. At each transition step of the Markov chain, the sentiment and aspect of the ith sentence are chosen according to the conditional probability

$$P\left(s_i = j, z_i = k | \mathbf{s}_{-i}, \mathbf{z}_{-i}, \mathbf{w}\right) \propto \frac{C_{dj}^{DS} + \gamma_j}{\sum_{j'=1}^{S} C_{dj'}^{DS} + \gamma_{j'}} \frac{C_{djk}^{DST} + \alpha_k}{\sum_{k'=1}^{T} C_{djk'}^{DST} + \alpha_{k'}}$$

$$\frac{\Gamma(\sum_{w=1}^{W} C_{jkw}^{STW} + \beta_{jw})}{\Gamma(\sum_{w=1}^{W} (C_{jkw}^{STW} + \beta_{jw}) + m_i)} \prod_{w=1}^{W} \frac{\Gamma(C_{jkw}^{STW} + \beta_{jw} + m_{iw})}{\Gamma(C_{jkw}^{STW} + \beta_{jw})}.$$

The notations are described in Table 1, with a minor exceptional use of notation that C_{dj}^{DS}, C_{djk}^{DST}, and C_{jkw}^{STW} in this expression exclude the sentence i.

The approximate probability of sentiment j in review d is

$$\pi_{dj} = \frac{C_{dj}^{DS} + \gamma_j}{\sum_{j'=1}^{S} C_{dj'}^{DS} + \gamma_{j'}}. \quad (1)$$

The approximate probability of aspect k for sentiment j in review d is

$$\theta_{djk} = \frac{C_{djk}^{DST} + \alpha_{jk}}{\sum_{k'=1}^{T} C_{djk'}^{DST} + \alpha_{jk'}}.$$

The approximate probability of word w in senti-aspect $\{k, j\}$ is

$$\phi_{jkw} = \frac{C_{jkw}^{STW} + \beta_{jw}}{\sum_{w'=1}^{V} C_{jkw'}^{STW} + \beta_{jw'}}.$$

5. EXPERIMENTAL SETUP

In this section, we describe our data sets and the sentiment seed words.

5.1 Data Sets

We use two different sets of reviews[1]. One data set is a collection of electronic device reviews from Amazon[2], which we name ELECTRONICS, and the other data set is restaurant reviews from Yelp[3], which we name RESTAURANTS. For ELECTRONICS, we collected all reviews in seven categories: air conditioner, canister vacuum, coffee machine, digital SLR, laptop, MP3 player, and space heater. We randomly selected at most 5,000 reviews from each category for balance, which resulted in about 22,000 total reviews. For RESTAURANTS,

---

2. For each document d,

   (a) Draw the document's sentiment distribution $\pi_d \sim$ Dirichlet(γ)

   (b) For each sentiment s, draw an aspect distribution $\theta_{ds} \sim$ Dirichlet(α)

   (c) For each sentence,
      i. Choose a sentiment $j \sim$ Multinomial(π_d)
      ii. Given sentiment j, choose an aspect k $\sim$ Multinomial(θ_{dj})
      iii. Generate words $w \sim$ Multinomial(ϕ_{jk})

ASUM exploits prior sentiment information by using asymmetric β. For example, we expect that the words "good, great" are not probable in negative expressions, and similarly the words "bad, annoying" are not probable in positive expressions. This can be encoded into β such that the elements of β corresponding to general positive sentiment words have small values for negative senti-aspects, and general negative sentiment words for positive senti-aspects. From the inference perspective, this asymmetric setting of β leads the words that co-occur with the general sentiment words to

---

[1]Available at http://uilab.kaist.ac.kr/research/WSDM11
[2]http://www.amazon.com
[3]http://www.yelp.com

Table 3: Full list of sentiment seed words in PARADIGM (bold) and PARADIGM+ (all). The first row is the positive words, and the second row is the negative words. The words' order is meaningless.

good, nice, excellent, positive, fortunate, correct, superior, amazing, attractive, awesome, best, comfortable, enjoy, fantastic, favorite, fun, glad, great, happy, impressive, love, perfect, recommend, satisfied, thank, worth
bad, nasty, poor, negative, unfortunate, wrong, inferior, annoying, complain, disappointed, hate, junk, mess, not_good, not_like, not_recommend, not_worth, problem, regret, sorry, terrible, trouble, unacceptable, upset, waste, worst, worthless

we collected the reviews of the 320 most rated restaurants in four cities: Atlanta, Chicago, Los Angeles, and New York City. We randomly selected 30,000 reviews out of the collected reviews.

We pre-processed the data by removing Web URLs and separating sentences by ".", "?", "!", and "newline". We removed words that contain non-English alphabets and sentences that are longer than 50 words. We used the Porter stemmer[4] for stemming. The properties of the data sets are summarized in Table 2.

Negation is an important issue in sentiment analysis, especially with the bag-of-words features. For example, in a sentence "the quality is not good", "not good" expresses negative sentiment, but without considering "not" and "good" collectively, it is hard to capture the negative sentiment. Previous work has proposed several approaches to this problem, including flipping the sentiment of a word when the word is located closely behind "not" [9]. We use simple regular expression rules to prefix "not" to a word that is modified by negating words, as was done in [8].

5.2 Sentiment Seed Words

We carefully chose affective words and general evalutive words for sentiment seed words. The seed words should not be aspect-specific evaluative words because they are assumed to be unknown. We use two sets of seed words. PARADIGM is the sentiment oriental paradigm words from Turney's work [24], containing seven positive words and seven negative words. PARADIGM+ is Turney's paradigm words plus other affective words and general evaluative words. The full list of the seed words is in Table 3.

6. EXPERIMENTS

We performed four experiments to evaluate our models, SLDA and ASUM. In the first experiment, we evaluate the aspects discovered by SLDA, and in the second experiment, we evaluate the senti-aspects discovered by ASUM. In the third experiment, we evaluate the sentiment words found by ASUM, and in the last experiment, we test the sentiment classification performance of ASUM.

It is worth noting that the aspects and senti-aspects computed by SLDA and ASUM are language models that are based on the word frequencies in the corpus. Hence, some words that are frequently used regardless of aspects may take top positions of the (senti-)aspects. To make characteristic words more apparent, we computed the term scores

---

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

---

on the discovered (senti-)aspects [21], which gives a lower score to the words common across various (senti-)aspects and higher score to the words that occur exceptionally often in one (senti-)aspect. All the aspects and senti-aspects shown in this section are based on the term scores, instead of the original probabilities calculated by the models.

6.1 Aspect Discovery

The first experiment is to automatically discover aspects in reviews using SLDA. We set three criteria for measuring the quality of the aspects. First, the discovered aspects should be coherent. Second, the aspects should be specific enough to capture the details in the reviews. Third, the aspects should be those that are discussed the most in the reviews. We applied SLDA to ELECTRONICS and RESTAURANTS data sets and evaluated the modeling power of SLDA based on these criteria. We also compared the results with LDA to see the effect of our assumption that one sentence represents one aspect. We varied the number of aspects and found that 50 aspects per sentiment captures various aspects with few redundancies, which we used for all the experiments in this section. We also tried several values of α and β but found that they do not really affect the quality of the result, and we use symmetric α and β set to be 0.1 and 0.001, respectively. Some examples of the discovered aspects are presented in Table 4.

From ELECTRONICS, SLDA discovered aspects that are specific to the seven product categories as well as general aspects such as design, orders, and service. SLDA discovered seven aspects about laptops–OS, MacBook, peripherals, battery life, hardware, graphics, and screen–as shown in Table 4(a). Each aspect represents a specific detail of the laptop. The aspects also cover most of the important parts and features of the laptop that users often point out and discuss in laptop reviews. These aspects are representative of the 50 aspects found, most of which are closely related to the product categories. These coherent, specific, and important aspects that SLDA found would be effective for potential applications such as aspect-level sentiment summarization and retrieval.

We compared the results of SLDA with the aspects found by LDA, and Table 5 presents the aspects related to cameras found by SLDA and LDA. For SLDA, we selected only five aspects out of 10 for space reasons. Both SLDA and LDA discovered the aspects "lens" and "iso". However, LDA could not find the aspects such as "grip", "beginners", and "ease of learning". These aspects are specific details that people evaluate about a camera. Overall, the aspects found by LDA tend to be more general and less coherent. This difference stems from our assumption built into SLDA that a single sentence represents one aspect. Accordingly, the aspects discovered by SLDA tend to account for the local positions of the words, which is an appropriate property for our goal. In contrast, LDA has a broader view that an aspect can be composed of any words in a review regardless of intra-sential word co-occurrences.

For RESTAURANTS, many of the SLDA aspects are related to cuisine types such as Mexican, seafood, breakfast, and dessert. The rest include parking, waiting, evaluation, and other general aspects about restaurants. Examples are presented in Table 4(b). The aspects "parking" and "waiting" are two detailed points that people often describe in restaurant reviews. LDA also discovers similar aspects except for

Table 4: Example aspects discovered by SLDA.

(a) ELECTRONICS

window	macbook	keyboard	batteri	laptop	usb	screen
vista	pro	pad	life	ram	port	bright
softwar	laptop	button	hour	processor	hdmi	displai
mac	appl	kei	charg	graphic	drive	color
instal	inch	mous	last	netbook	connect	glossi
os	screen	touch	recharg	drive	dvd	lcd
xp	mac	trackpad	power	core	wireless	keyboard
run	aluminum	finger	charger	game	video	reflect
program	unibodi	touchpad	cell	batteri	card	light
driver	new	scroll	long	hp	extern	angl
pc	displai	click	hr	notebook	tv	glare
comput	keyboard	screen	laptop	gb	movi	view
microsoft	trackpad	type	mode	comput	laptop	led
boot	mbp	laptop	run	intel	cabl	clear
laptop	glossi	gestur	fulli	screen	speaker	inch

(b) RESTAURANTS

park	wait	beer	yum
street	line	wine	oh
valet	seat	drink	no
cash	crowd	glass	mmm
lot	long	select	mmmm
meter	weekend	bottl	ye
across	get	martini	wow
car	tabl	tap	love
find	reserv	mojito	yeah
free	your	margarita	lol
block	earli	cocktail	holi
onli	hour	sangria	haha
valid	if	juic	omg
there	there	list	not
walk	busi	vodka	yuck

Table 5: Discovered aspects regarding cameras for SLDA and LDA.

(a) SLDA

camera	iso	len	camera	camera
hand	card	kit	dslr	learn
feel	raw	lens	slr	easi
grip	imag	zoom	photographi	manual
weight	camera	af	photograph	menu
size	shoot	camera	rebel	mode
fit	nois	canon	digit	set
solid	photo	ef	shoot	control
small	file	nikon	canon	shoot
bodi	print	vr	beginn	featur
rebel	pictur	nikkor	amateur	intuit
light	jpeg	dx	recommend	auto
comfort	shot	flash	profession	pictur
batteri	resolut	usm	nikon	user
smaller	memori	bodi	point	photographi

(b) LDA

nikon	light	flash	camera	camera	camera
len	iso	camera	pictur	canon	canon
lens	color	card	shoot	len	shoot
pentax	imag	batteri	photo	digit	focu
olympu	nois	memori	point	rebel	shot
qualiti	low	digit	shot	slr	view
imag	high	shot	digit	pictur	imag
dslr	set	set	learn	lens	iso
bodi	qualiti	sd	great	kit	lcd
af	bright	pictur	manual	nikon	pictur
kit	nikon	speed	slr	xt	mode
soni	raw	fast	photographi	qualiti	auto
focu	white	mode	photograph	xti	len
zoom	perform	shoot	set	film	live
system	balanc	second	nikon	bodi	sensor

the last two aspects in the table, "liquors" and "interjections". In the LDA result, the top words in "liquors" are spread out across different aspects. For example, "beer" appears in an aspect related to bars, and "wine" appears in an aspect related to desserts. This shows that LDA captures more global aspects from reviews. When we look at the interjections, they also appear across various aspects for LDA. Without considering sentence boundaries, the interjections didn't have enough evidence to be formed into one aspect. In SLDA, on the other hand, the co-occurrences of these words within sentence boundaries cause them to form an aspect. Interjections may be hard to be considered as an "aspect". Yet, they play an important role of expressing sentiment in restaurant reviews, and knowing the usages of those words in the corpus and in each of the reviews leads to a better understanding of the reviews.

6.2 Senti-Aspect Discovery

Our second experiment is to discover *senti-aspects*, aspects coupled with a *sentiment* (positive or negative). For example, the "screen" aspect discovered by SLDA (Table 4(a)) contains the words "screen, bright, displai, color, light, lcd, look, like, glossi", whereas a {screen, negative} senti-aspect discovered by ASUM (Table 6(a)) contains the words "screen, glossi, glare, reflect". We can apply the same criteria to evaluate senti-aspects as the criteria in Section 6.1 for aspects. Additionally, each senti-aspect should clearly

represent its sentiment. We applied ASUM to ELECTRONICS and RESTAURANTS. For both data sets, the number of aspects is set to be 70 for each sentiment.

ASUM needs two hyperparameters, γ and β, to be tuned carefully. γ is a prior for the sentiment distribution in a review. Because we assume no prior knowledge of the sentiment distribution, we simply use a symmetric γ of 1, which means all possible sentiment distributions are equally likely.

The second hyperparameter, β, is one of two key elements for integrating the sentiment seed words into ASUM, and the other key element is the initialization of Gibbs sampling. Both of these elements must be carefully chosen for the seed words to be effective. β is the prior of the word distributions of senti-aspects, and we use asymmetric β for ASUM. For positive senti-aspects, we set the elements of β to be 0 for the negative seed words and 0.001 for all the other words. Similarly, for negative senti-aspects, we set the elements of β to be 0 for the positive seed words and 0.001 for all the other words. This indicates that we initially predict that no negative seed word appears in positive senti-aspects, and vice versa. With these asymmetric priors, if we use a random initialization of Gibbs sampling, as it is usually done, the asymmetric priors would just be ignored. Therefore, in the initialization step, we assign the sentiment seed words their seed sentiment. We chose this setting because it is effective and simple, but the limitation is that the sentiment seed words can only be assigned to the senti-aspects of the same

Table 6: Example senti-aspects discovered by ASUM. The labels are manually annotated.

(a) ELECTRONICS

price(p)	price(n)	screen(p)	screen(n)	screen(n)	screen(n)	vacuum(p)
worth	monei	screen	screen	fingerprint	screen	easi
monei	save	color	glossi	glossi	font	light
penni	notwast	bright	displai	magnet	point	carri
extra	wast	clear	keyboard	screen	size	weight
well	yourself	video	bright	show	notebook	lightweight
everi	notbui	displai	glare	finger	ssd	suction
price	awai	crisp	angl	finish	kei	small
dollar	spend	great	view	print	shoot	around
spend	notworth	resolut	color	smudg	smaller	vacuum
pai	stai	qualiti	light	easili	pictur	power
save	time	pictur	lcd	scratch	sensor	stair
cost	favor	sound	reflect	reflect	small	compact
hundr	pleas	sharp	matt	cover	sens	quiet
buck	heater	movi	edg	dust	lol	move
definit	headach	beauti	macbook	prone	photo	handl

(b) RESTAURANTS

meat(p)	meat(n)	music(p)	music(n)	interjection(p)	interjection(n)	payment(n)
flavor	dry	music	loud	mouth	yuck	cash
tender	bland	night	tabl	mmm	sigh	onli
crispi	too	group	convers	wow	digress	card
sauc	salti	crowd	hear	melt	meh	credit
meat	tast	loud	music	omg	wtf	downsid
juici	flavor	bar	nois	good	boo	park
soft	meat	atmospher	talk	holi	yai	take
perfectli	chicken	peopl	sit	nom	mmmmmm	accept
veri	bit	dinner	close	water	dunno	bring
moist	littl	fun	other	yummi	bummer	wait
sweet	pork	good	each	yum	wow	dun
perfect	sauc	great	room	oh	notcool	neg
cook	lack	date	can	mmmmm	bleh	complaint
crust	chewi	go	space	delici	haha	lack
fresh	disappoint	plai	peopl	serious	hoorai	make

sentiment. We can loosen the limitation by using different values for β and different initialization, and we leave this for future work. For the experiments in this section, we used PARADIGM+ as the sentiment seed words. Examples of the senti-aspects that ASUM discovered are in Table 6.

The senti-aspects found by ASUM from the ELECTRONICS data set illustrates how the consideration of sentiment affects the discovered aspects. While SLDA tends to find aspects according to the product categories (laptop, mp3 player, DSLR camera), ASUM finds some aspects that span various product categories while invoking similar sentiments. An example is the "screen" aspect, shown in Table 4(a) compared to the "screen(p)" and "screen(n)" senti-aspects in Table 6(a). In the SLDA results, there is one aspect devoted to a general screen, and there are a few aspects about the screen of specific products, such as the "macbook" aspect and an "mp3 player" aspect (not shown) that are about product features including the screen. In the ASUM results, there is a "screen(p)" senti-aspect that represents the general positive sentiment about screens, and then there are three "screen(n)" senti-aspects, each about a different reason that might invoke a negative sentiment toward screens for various product categories. We can infer from these senti-aspects that users expressed negative sentiments because of the "glare of the glossy screen", "fingerprints easily left on the screen", and "small size of the screen". The joint modeling of the sentiment and the aspect in ASUM gives rise to this different behavior, which would enable understanding

of users' sentiments at the level of common features across various product categories. The first two senti-aspects in the table show how ASUM aligns relevant sentiment words with the sentiment seed words for an aspect. For example, in "price(p)", the seed word "worth" is aligned with other sentiment words such as "extra, well, save". In "price(n)", the seed word "not_worth" is aligned with the words "not_waste, not_buy", which are used in reviews as "Do not waste your money on this." The word "save" is probable in both positive sentiment and negative sentiment. It is because one word can indicate different sentiments depending on the syntax. A word may invoke different sentiments depending on the context or aspect as well. The "vacuum(p)" senti-aspect in the table is about portability. In this senti-aspect, the word "small" is used positively, whereas "small" is used negatively in "screen(n)". This shows the power of our probabilistic approach that a word is not limited to one sentiment.

Results from running ASUM on RESTAURANTS show that ASUM can discover senti-aspects for which only one sentiment is present in the corpus. The example of "payment", shown in Table 6(b), only exists for the negative sentiment, represented by the words "cash, only, card, accept" describing the negative sentiment of the users with the cash only policy of a restaurant. The RESTAURANTS results also confirm that there are some aspects that surface with the interplay between sentiment and aspects, as we can see from the "meat" aspect. For example, to express positive sentiment on "meat", people use words like "tender", "crispy", "juicy", and

Table 7: Automatically detected sentiment words. The senti-aspects discovered by ASUM were utilized to illustrate different sentiment words for the same aspect.

Common Words	Sentiment Words
screen color bright displai crisp qualiti sharp	clear great pictur sound movi beauti good hd imag size watch rai nice crystal
	glossi glare light reflect matt edg macbook kei black bit peopl notlik minor
music song player video download itun zune file	radio listen fm movi record easi convert podcast album audio book librari watch
	problem updat driver vista system xp firmwar disk mac hard run microsoft appl
our us server waiter tabl she he waitress ask minut seat	water glass refil wine attent friendli brought sat veri arriv plate help staff nice
	said me want card get tell if would gui bad could rude pai becaus walk then

Table 8: Sentiment classification by the generative models and the supervised classifiers. The number of aspects is 70 for each sentiment.

	ELECTRONICS	RESTAURANTS
Baseline	0.81	0.85
LingPipe-Uni	0.71	0.81
LingPipe-Bi	0.79	0.87
ASUM	**0.78**	**0.79**
ASUM+	**0.84**	**0.86**
JST+	0.65	0.60
TSM+	0.48	0.52

"crust". To express negative sentiment, they use words such as "dry", "bland", and "disappointed". These two aspects were discovered in ASUM but not in SLDA, and the reason is that people express their sentiment toward these aspects very clearly. In SLDA the words that convey a sentiment toward the quality of meat appear in various cuisine-type aspects such as steak, burger, and pizza. Because people often evaluate specifically on the quality of meat, however, these words become apparent in ASUM.

6.3 Aspect-Specific Sentiment Words

The joint modeling of aspect and sentiment means ASUM finds, as top probability words in each of the senti-aspects, both aspect words, and sentiment words that are dependent on the aspect. Since we start with a set of general sentiment words, this yields the effect of bootstrapping the general sentiment words to discover aspect-specific sentiment words. This is one advantage of ASUM over TSM [16], in which all topics share one sentiment word distribution.

We introduce a simple method for employing the result of ASUM to automatically distinguish between positive and negative sentiment words for the same aspect. This increases the utility of ASUM by providing an organized result that shows why people express sentiment toward an aspect and what words they use. The process is as follows:

1. Calculate the cosine similarity between every pair of senti-aspects with different sentiments.

2. If the similarity exceeds a certain threshold, two senti-aspects are considered to represent the same aspect.

3. If a word takes a high probability in both senti-aspects, then this word is a common word.

4. If a word takes a high probability in only one senti-aspect, then this word is a sentiment word whose sentiment follows the senti-aspect.

We applied this method to our data sets and present the results in Table 7. For a music player, people praised the converting process, but they did not like driver and firmware updates. In the restaurant reviews, people praised waiters and waitresses for being attentive and friendly, but they complained when the servers were rude. The overall results show that ASUM discovers aspect-specific sentiment words,

which can be used in applications such as review summarization.

6.4 Sentiment Classification

In this section, we present the results of sentiment classification to quantitatively evaluate the quality of senti-aspects discovered by ASUM. To determine the sentiment of a review, we use π (Equation 1), the probabilistic sentiment distribution in a review, such that a review is set to be positive if positive sentiment has the equal or a higher probability than negative sentiment, and set to be negative otherwise. Both ELECTRONICS and RESTAURANTS use the 5-star rating system, and the ratings of 1 or 2-stars are treated as negative and 4 or 5-stars positive. We do not classify on the reviews with 3-stars, but they are still used to fit ASUM to the data. The hyperparameters of ASUM are set to be the same as in the experiments in Section 6.2.

We compare the performance of ASUM with JST [15], TSM [16], LingPipe [1] (Unigrams & Bigrams), and the baseline. LingPipe first separates subjective sentences from objective sentences, and then finds sentiment using word features. The baseline classifies each review according to the numbers of sentences that contain the positive and negative sentiment seed words.

The classification results are presented in Figure 3 in terms of accuracy. The baseline and LingPipe are not shown in the figure because of space, but they are shown numerically in Table 8. In all settings, ASUM outperforms the other unsupervised models and even supervised LingPipe in the same condition of unigrams. The baseline with only the seed words performs quite well, but ASUM performs even better. In general, the accuracy increases as the number of aspects increases because the models better fit the data. JST had great performance on movie reviews in the original paper [15], but did not perform well on our data. TSM is not intended for sentiment classification, and sentiment words are not adapted to aspects. In the original paper [16], TSM was used to analyze topic life cycles and sentiment dynamics.

The visualization of the assignment of senti-aspects to each sentence would help to understand and analyze the reviews. The posterior probability of the senti-aspects for each sentence was used to visualize reviews. Two examples are shown in Figure 4. The visualization shows that the sentiments were found to be quite accurate. It is worth noting that sentences that are too short are difficult to assign correct senti-aspects because they may lack strong evidence for sentiment and aspects.
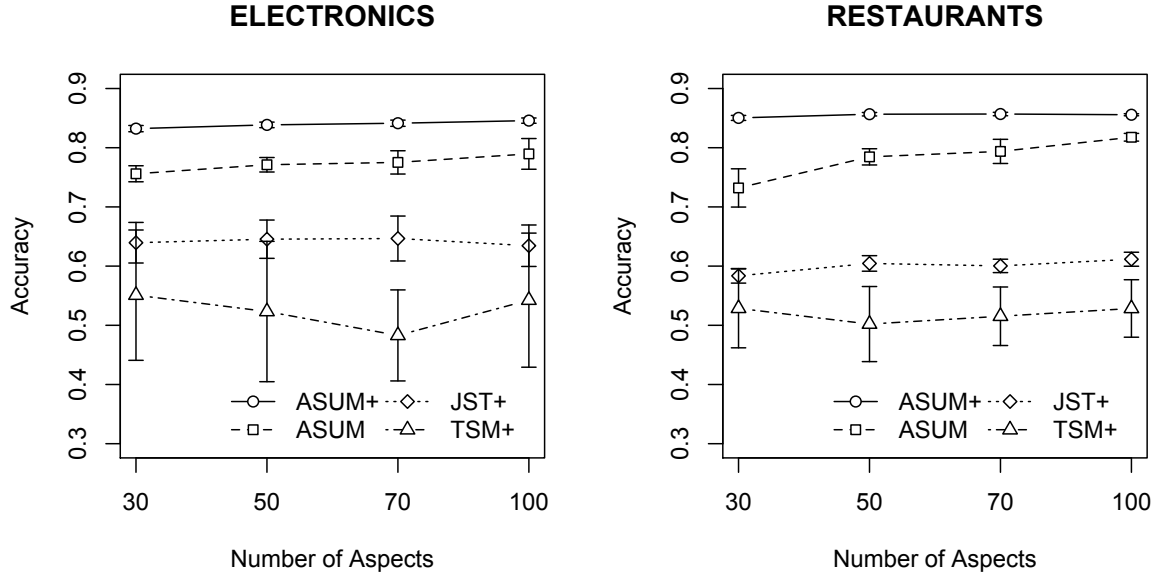
ELECTRONICS   **RESTAURANTS**

Figure 3: Sentiment classification by the three unified models. "ASUM" uses PARADIGM, and "ASUM+", "JST+", and "TSM+" use PARADIGM+. The error bars represent the standard deviation from 10 samples.

{-, 61}	When my old maker broke, I tried several (at family and friends houses, bought and returned one), then decided on the steel Cuisinart.
{-, 61}	The coffee just wasn't as good as my old Krups.
{+, 56}	Bought the FME2 (just like the 4) and it has made amazing coffee for almost three years.
{+, 58}	Better than starbucks by a mile, as good as Peets.
{+, 66}	This one is a winner.
{-, 39}	The earlier negative reviews seem based on nothing but incompetence, if you actually read them.

(a) ELECTRONICS

{+, 52}	The restaurant is really pretty inside and everyone who works there looks like they like it.
{+, 32}	The food is really great.
{+, 17}	I would recommend any of their seafood dishes.
{+, 33}	Come during happy hour for some great deals.
{-, 55}	The reason they aren't getting five stars is because of their parking situation.
{-, 57}	They technically don't "make" you use the valet but there's only a half dozen spots available to the immediate left.

(b) RESTAURANTS

Figure 4: Visualization of reviews. The first column is the senti-aspects of the sentences. (+: positive, −: nagetive)

7. CONCLUSION

In this paper, we proposed two generative models to discover aspects and sentiment in reviews. SLDA constrains that all words in a single sentence be drawn from one aspect. ASUM unifies aspects and sentiment and discovers pairs of {aspect, sentiment}, which we call senti-aspects. The aspects and senti-aspects discovered from reviews of electronic devices and restaurants show that SLDA and ASUM capture important evaluative details of the reviews. ASUM is also capable of capturing aspects that are closely coupled with a sentiment. We showed that the senti-aspects found by ASUM can be used to illustrate different sentiments toward the same aspect, which would be utilized in applications such as review summarization. In the quantitative evaluation of sentiment classification, ASUM outperformed other generative models and came close to supervised classification methods.

There are several possibilities to improve our models. Since our models assume that one sentence contains exactly one aspect, we may split sentences not only by punctuations but also by conjunctions. In addition, we may use a part-of-speech tagger for more accurate negation detection [6]. To make each (senti-)aspect clearer, we can filter out words that are common across many (senti-)aspects by adding a background language model.

For future work, our models may be utilized for aspect-based review summarization. We can apply the models to other types of data such as editorials and art critiques, or use different seed words to capture different dimensions than sentiment.

8. ACKNOWLEDGMENTS

9. REFERENCES

[1] Alias-i. Lingpipe 4.0.1. http://alias-i.com/lingpipe, 2008.
[2] A. Aue and M. Gamon. Customizing sentiment classifiers to new domains: A case study. In *Proceedings of Recent Advances in Natural Language Processing (RANLP)*, 2005.

[3] 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)*, New York, NY, USA, 2008. ACM.

[4] D. M. Blei, A. Y. Ng, M. I. Jordan, and J. Lafferty. Latent dirichlet allocation. *Journal of Machine Learning Research*, 3:993–1022, 2003.

[5] J. Blitzer, M. Dredze, and F. Pereira. Biographies, bollywood, boom-boxes and blenders: Domain adaptation for sentiment classification. In *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics*, pages 440–447, Prague, Czech Republic, June 2007. Association for Computational Linguistics.

[6] 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*, pages 804–812, Los Angeles, California, June 2010. Association for Computational Linguistics.

[7] Y. Choi, Y. Kim, and S.-H. Myaeng. Domain-specific sentiment analysis using contextual feature generation. In *Proceeding of the 1st international CIKM workshop on Topic-sentiment analysis for mass opinion*, pages 37–44, New York, NY, USA, 2009. ACM.

[8] S. R. Das, M. Y. Chen, and Efa. Yahoo! for amazon: Sentiment parsing from small talk on the web. *Social Science Research Network Working Paper Series*, August 2001.

[9] K. Eguchi and V. Lavrenko. Sentiment retrieval using generative models. In *Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing*, pages 345–354, Sydney, Australia, July 2006. Association for Computational Linguistics.

[10] T. L. Griffiths and M. Steyvers. Finding scientific topics. *Proceedings of the National Academy of Sciences of the United States of America*, 101(Suppl 1):5228–5235, 2004.

[11] T. Hofmann. Probabilistic latent semantic indexing. In *Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval*, pages 50–57, New York, NY, USA, 1999. ACM.

[12] 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*, pages 168–177, New York, NY, USA, 2004. ACM.

[13] V. Jijkoun, M. de Rijke, and W. Weerkamp. Generating focused topic-specific sentiment lexicons. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, pages 585–594, Uppsala, Sweden, July 2010. Association for Computational Linguistics.

[14] N. Kaji and M. Kitsuregawa. Building lexicon for sentiment analysis from massive collection of HTML documents. In *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL)*, pages 1075–1083, Prague, Czech Republic, June 2007. Association for Computational Linguistics.

[15] C. Lin and Y. He. Joint sentiment/topic model for sentiment analysis. In *Proceeding of the 18th ACM conference on Information and knowledge management*, pages 375–384, New York, NY, USA, 2009. ACM.

[16] 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*, pages 171–180, New York, NY, USA, 2007. ACM.

[17] S. Mohammad, C. Dunne, and B. Dorr. Generating high-coverage semantic orientation lexicons from overtly marked words and a thesaurus. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing*, pages 599–608, Singapore, August 2009. Association for Computational Linguistics.

[18] S. J. Pan, X. Ni, J.-T. Sun, Q. Yang, and Z. Chen. Cross-domain sentiment classification via spectral feature alignment. In *Proceedings of the 19th international conference on World wide web*, pages 751–760, New York, NY, USA, 2010. ACM.

[19] 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*, pages 339–346, Morristown, NJ, USA, 2005. Association for Computational Linguistics.

[20] D. Rao and D. Ravichandran. Semi-supervised polarity lexicon induction. In *Proceedings of the 12th Conference of the European Chapter of the ACL (EACL 2009)*, pages 675–682, Athens, Greece, March 2009. Association for Computational Linguistics.

[21] A. Srivastava and M. Sahami, editors. *Text Mining: Classification, Clustering, and Applications*, chapter 4, page 75. CRC Press, Boca Raton, FL, 2009.

[22] I. Titov and R. Mcdonald. A joint model of text and aspect ratings for sentiment summarization. In *Proc. ACL-08: HLT*, pages 308–316, 2008.

[23] I. Titov and R. McDonald. Modeling online reviews with multi-grain topic models. In *Proceeding of the 17th international conference on World Wide Web*, pages 111–120, New York, NY, USA, 2008. ACM.

[24] P. D. Turney and M. L. Littman. Measuring praise and criticism: Inference of semantic orientation from association. *ACM Transactions on Information Systems*, 21(4):315–346, 2003.

[25] H. M. Wallach. Topic modeling: beyond bag-of-words. In *Proceedings of the 23rd international conference on Machine learning*, pages 977–984, New York, NY, USA, 2006. ACM.

[26] X. Zhao, J. Jiang, H. Yan, and X. Li. Jointly modeling aspects and opinions with a MaxEnt-LDA hybrid. In *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*, pages 56–65, Cambridge, MA, October 2010. Association for Computational Linguistics.