Aspect and Sentiment Unification Model

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 in online markets, blogs, and communities contain sentiments about detailed aspects of the products and services reviewed. However, most of the reviews are plain text and thus require much effort to obtain information about relevant details. In this paper, we present the *Aspect and Sentiment Unification Model* (ASUM), a probabilistic generative model, to automatically discover aspects people evaluate and different sentiments toward these aspects. ASUM incorporates sentiment and aspects together to discover from reviews the aspects that are evaluated positively and the ones evaluated negatively. We applied the model to reviews of electronic devices and restaurants and photo critiques. The results show that the aspects discovered by ASUM match evaluative details of the reviews and capture important aspects that are closely coupled with a sentiment. On sentiment classification tasks, ASUM outperforms other generative models and comes close to supervised classification methods even though this model does not use the sentiment labels of the reviews at all.

1   Introduction

The Web has an overwhelming amount of user-generated contents about their experiences of products, restaurants, photos, etc. In those contents, people contribute to the information by praising and criticizing a variety of *aspects* of the target of the review, such as the noise level of a vacuum cleaner or the waiting time of a restaurant. Although some Websites (e.g., TripAdvisor) are specifically designed for user reviews with a predefined evaluation form, most of the users express their opinions in online communities and personal blogs using free text without any structure.

From the perspective of a user reading the reviews to get information from other users who have already had similar experiences, the evaluations of specific aspects are just as important as the overall rating. A user looking to buy a digital camera may want to know what a review says about the photo quality and the brightness of the lens, not just whether the review recommends the camera. From the perspective of engines designed to automatically retrieve contents related to a specific experience, it is also important to know how the opinions and sentiments for different aspects of the experience are expressed. 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., electronic devices, movies, restaurants).

We present the *Aspect and Sentiment Unification Model* (ASUM), a probabilistic generative topic model, to tackle these two problems at once. This model discovers pairs of {sentiment, aspect}, which we call *senti-aspects*, in an unsupervised way. A senti-aspect is interpreted as an aspect that is evaluated with a specific sentiment, and it is represented as a probability distribution over words, which explains how much a word is related to the aspect and sentiment. An example senti-aspect can be "good portability," which may be closely related with the words light, small, carry, and easy.

A widely used approach in sentiment analysis utilizes part-of-speech information to extract aspect and sentiment words [2]. In some work, a predefined set of aspects are given [8]. These approaches are not flexible or comprehensive enough. Several topic modeling approaches have been proposed as well [4, 5, 6]. They jointly model topics and sentiment, describing generative processes of opinionated documents. These models, however, are not appropriate to find evaluative details and aspect-specific sentiment words.
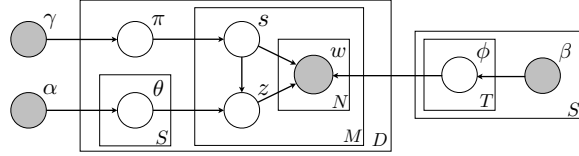
2 Model

Figure 1: Graphical representation of ASUM

ASUM models the generative process of a document as illustrated in the following scenario of writing a review. There are a fixed number of positive aspects and negative aspects. 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 distributions of the positive aspects and negative aspects, 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. The graphical representation of ASUM is shown in Figure 2. Formally, the generative process is as follows:

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

2. For each document d,

    (a) Draw the document's sentiment distribution $\pi_d \sim \text{Dirichlet}(\gamma)$

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

    (c) For each sentence,

        i. Choose a sentiment $j \sim \text{Multinomial}(\pi_d)$
        ii. Given sentiment j, choose an aspect $k \sim \text{Multinomial}(\theta_{dj})$
        iii. Given sentiment j and aspect k, generate every word $w \sim \text{Multinomial}(\phi_{jk})$

ASUM exploits prior sentiment information by using asymmetric β. Considering some sentiment seed words that express a consistent sentiment every time they are used, for example, `good`, `satisfied`, `bad`, and `annoying`, we expect that the positive words are not probable in negative expressions, and similarly the negative words are not probable in positive expressions. It can be encoded into β such that the elements of β corresponding to the positive seed words have small values for negative senti-aspects, and vice versa. From the inference perspective, this asymmetric setting of β leads the words that co-occur with the seed words to be more probable in the corresponding sentiment.

Latent variables θ, π, and ϕ are inferred by Gibbs sampling. Suppose there are S sentiments, T aspects for each sentiment, and W unique words. At each transition step of the Markov chain, for sentence i, sentiment j and aspect k are chosen according to the conditional probability

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

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

where $\mathbf{s}_{-i}$ and $\mathbf{z}_{-i}$ indicate the aspect and sentiment assignments respectively, both excluding sentence i. C_{dj}^{DS} is the number of times sentiment j has occurred in document d, C_{djk}^{DST} is the number

Table 1: Full list of sentiment seed words in PARADIGM (bold) and PARADIGM+ (all)

Positive	**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
Negative	**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

of times sentiment s and aspect k together occurred in document d, and C_{jkw}^{STW} is the number of times word w is assigned to sentiment j and aspect k. All these counters exclude sentence i. m_i is the total number of words in sentence i, and m_{iw} is the frequency of word w in sentence i.

The probability of sentiment j in review d, the probability of aspect k with sentiment j in review d, and the probability of word w in senti-aspect $\{j,k\}$ are approximated as

$$\pi_{dj} = \frac{C_{dj}^{DS} + \gamma_j}{\sum_{j'=1}^{S} C_{dj'}^{DS} + \gamma_{j'}}, \theta_{djk} = \frac{C_{djk}^{DST} + \alpha_{jk}}{\sum_{k'=1}^{T} C_{djk'}^{DST} + \alpha_{jk'}}, \phi_{jkw} = \frac{C_{jkw}^{STW} + \beta_{jw}}{\sum_{w'=1}^{V} C_{jkw'}^{STW} + \beta_{jw'}},$$

now C^{DS}, C^{DST}, and C^{STW} including all sentences.

3 Experiments

We use three review sets to see how ASUM behaves in different domains.

- **ELECTRONICS** 24,000 reviews from Amazon.com that fall into the categories *air conditioner, canister vacuum, coffee machine, digital SLR, laptop, MP3 player*, and *space heater*
- **RESTAURANTS** 27,000 reviews from Yelp.com that fall into four cities *Atlanta, Chicago, Los Angeles*, and *New York City*
- **PHOTOGRAPHY** 27,000 photo critiques from PhotoSIG.com

We removed stop words and used the Porter stemmer for stemming. We used a simple regular expression rule to prefix `not` to a negated word.

To incorporate sentiment information into ASUM, we use two sets of seed words. The first set PARADIGM is the sentiment oriental paradigm words from Turney's work [7], which contain seven positive words and seven negative words. For the second set PARADIGM+, we carefully added aspect-independent sentiment words to PARADIGM. The full list of the seed words is in Table 1.

Throghout the experiments we used two sentiments (positive/negative) and 70 aspects for each sentiment, and PARADIGM+ was used for sentiment seed words. We empirically set $\alpha = 0.1$ and

Table 2: Example senti-aspects discovered by ASUM. The labels are manually annotated.

(a) ELECTRONICS			(b) RESTAURANTS			(c) PHOTOGRAPHY	
heater(p)	heater(n)	heater(n)	payment(n)	meat(p)	meat(n)	greeting(p)	improve(n)
room	smell	nois	cash	flavor	dry	regard	more
heat	heater	fan	onli	tender	bland	best	if
heater	odor	loud	card	crispi	too	andr	better
warm	burn	quiet	credit	sauc	salti	gisela	could
bedroom	hour	sound	downsid	meat	tast	tom	think
small	first	noisi	park	juici	flavor	janet	littl
space	heat	heater	take	soft	meat	kevin	been
up	turn	hear	accept	perfectli	chicken	liux	crop
bathroom	oil	room	bring	veri	bit	fong	shot
keep	run	turn	wait	moist	littl	jeff	bit
feet	few	heat	dun	sweet	pork	harri	right
area	bad	sleep	neg	perfect	sauc	enjoi	angl

ELECTRONICS   RESTAURANTS
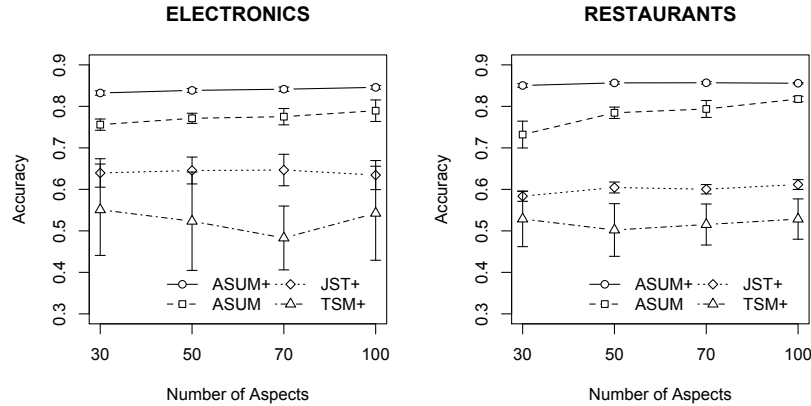
Figure 2: Sentiment classification results. "+" indicates PARADIGM+.

{-, 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.

{+, 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.

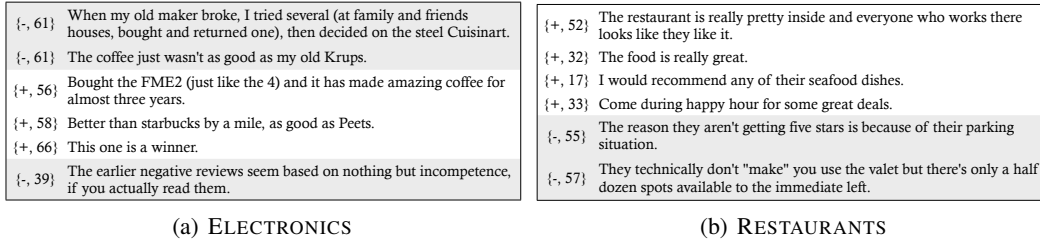(a) ELECTRONICS   (b) RESTAURANTS

Figure 3: Visualization of two reviews. The left columns are senti-aspects (+: positive, −: negative)

$\gamma = 1$. For positive(negative) senti-aspects, we set β to be 0 for the negative(positive) seed words and 0.001 for the other words. In the initialization step of Gibbs sampling, we assigned the sentiment seed words to their seed sentiments. By doing this, the sentiment seed words can occur only in the senti-aspects of their sentiment.

Example senti-aspects are shown in Table 2. Table 2(a) shows one positive aspect and two negative aspects about heaters, each about a different reason of sentiment toward heaters. When we compare the senti-aspects with LDA results [3], it can be seen that ASUM finds fine-grained evaluative aspects of one product category, whereas LDA tends to group several details into one topic. Table 2(b) shows a negative senti-aspect about payment. ASUM captured only negative senti-aspect about payment because people say about the cash-only policy negatively most of the time. ASUM found two senti-aspects about meat, but in LDA, sentiment words about the quality of meat appear in various cuisine-type topics. As people often evaluate specifically on the quality of meat, no matter what the food type is, the interplay between sentiment and aspect in ASUM captures these sentiment words as one senti-aspect. Table 2(c) shows one of the senti-aspects about greetings in PHOTOGRAPHY. They differ in how to start, e.g., cheers, wishes. Notice that the senti-aspect about how photos could be improved is discovered as a negative senti-aspect, reflecting the characteristic of the PHOTOGRAPHY data that critics tend to use kind and explanatory language.

We performed sentiment classification on ELECTRONICS and RESTAURANTS. The sentiment of a review is determined to be the sentiment that takes the higher probability in π. Both data sets use the 5-star rating system, so for the ground truth, 1 or 2-stars is treated as negative and 4 or 5-stars positive. We did not classify on the reviews with 3-stars, but they were still used to fit the model. We compared the classification performance of ASUM with other unsupervised joint models JST [4] and TSM [5] and with supervised classifiers provided by LingPipe (Unigrams & Bigrams) [1]. The results are presented in Figure 2 in terms of accuracy. LingPipe achieved the accuracies of 0.71 (Uni), 0.79 (Bi) for ELECTRONICS and 0.81 (Uni), 0.87 (Bi) for RESTAURANTS. ASUM outperforms the other unsupervised models and even supervised LingPipe in the same condition of unigrams. We visualized the sampling results in Figure 3. The visualization shows that the sentiments were found to be quite accurate.

4

References

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

[2] 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.

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

[4] 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.

[5] 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.

[6] 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.

[7] 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.

[8] H. Wang, Y. Lu, and C. Zhai. Latent aspect rating analysis on review text data: A rating regression approach. In *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 783–792, New York, NY, USA, 2010. ACM.