Similarity Dependency Dirichlet Process for Aspect-Based Sentiment Analysis

Wanying Ding* Xiaoli Song† Yue Shang‡ Junhuan Zhu§ Lifan Guo¶
Xiaohua Hu‖

Abstract

Aspect-base Sentiment Analysis is a core component in *Review Recommendation System*. With the booming of customers' reviews online, an efficient sentiment analysis algorithm will substantially enhance a review recommendation system's performance, providing users with more helpful and informative reviews. Recently, two kinds of LDA derived models, namely *Word Model* and *Phrase Model*, take the dominant positions in this field. However, the requirement of exact aspect number underlies the usability and flexibility of these LDA extended models. Although, Dirichlet Process(DP), which can help to automatically generate the number of aspects, has been on trial, its random word assignment mechanism makes the result unsatisfying. This paper proposes a model named Similarity Dependency Dirichlet Process(SDDP) to cope with the above problems. SDDP inherits the merits of DP to automatically determine the number of aspects, but it exploits the semantic similarities between words to infer aspects and sentiments, alleviating the random assignment problem in DP. Furthermore, based on SDDP, this paper builds a word model W-SDDP and a phrase model P-SDDP respectively to detect aspects and sentiments from two different perspectives. Finally we experiment both our two models on 6 datasets, and compare with other 6 currently most popular models. The result shows that both W-SDDP and P-SDDP outperform the other 6 models, indicating SDDP is a promising model for sentiment analysis.

1 Introduction

Social media has provided an open platform for users to exchange ideas online, and more and more valuable opinions have been published on different websites,like Amazon and Yelp. But with the dramatic increase in volume, it will cost customers dozens of hours going through all the reviews. Thus, *Review Recommendation System* has been created in order to present customers with most helpful or informative reviews. Within such a system, aspect-based sentiment analysis is an important component[1],[2] to make it function, because it can efficiently detect the ***Aspect*** *(A particular feature, like food or service of a restaurant)* and corresponding ***Sentiments*** *(The subjectivity polarities, like positive or negative)*,and further provide vital features to recognize helpful reviews.

As a core component of recommendation system, sentiment analysis has been long explored. In the early days, classification methods were widely applied [3],[4],[5],[6], and laid a solid foundation. However, the requirement of training dataset presents a challenge for related researches, since manually constructing training datasets is time consuming and laborious, but more and more online data presents as unlabelled. Thus, unsupervised methods become more attractive because of its label-free and flexible in use. The birth of Latent Dirichlet Allocation(LDA)[7] has inspired a lot of researchers in developing unsupervised models[8],[9],[10],[11],[12],[13],[14]. LDA extended models inherit both LDAs advantages and disadvantages. Speaking of advantages, they are all training dataset free, and efficient in aspect and sentiment discovery. The disadvantage is that they require users to decide the number of aspects beforehand. Such decisions are hard to make, especially when one has little knowledge about the dataset at hand. Although it is possible to repeatedly check the perplexity values to find an optimal number of aspects, it has a major practical limitation in that one may have to make a large number of trials. Recently, a few researchers try to replace the static Dirichlet allocation in LDA with dynamic Dirichlet process (DP), which can automatically generate the aspect number according datasets' own characteristics. Jianfeng Si[15] took the first try in implementing DP to discover the sentiment from Twitter, and Kim Suin[16]adopted the Hierarchical DP(HDP)[17] to explore the hierarchical aspect structure from online reviews. Both Jianfeng Si and Kim Suin's works are just simply to apply traditional DP, and ignore the random word assignment problem, which is emphasized by Blei

*College of Computing and Informatics,Drexel University.
†College of Computing and Informatics,Drexel University.
‡College of Computing and Informatics,Drexel University.
§Department of Computer Science, Rochester University.
¶TCL Research America.
‖College of Computing and Informatics,Drexel University.

and Frazier[18].

Considering all the problems mentioned above, this paper has proposed a Similarity Dependency Dirichlet Process (SDDP) for sentiment analysis. Comparing to LDA, SDDP can automatically determine the number of aspects, saving users from laborious trials. Comparing to DP or HDP, SDDP replaces the random assignment mechanism with semantic similarity assignment, making the model more reasonable and acceptable. Besides, considering the *Word Model* and the *Phrase Model* are two research logistics in sentiment analysis area, this paper builds two models based on SDDP, one word model (W-SDDP) and one phrase model (P-SDDP), to recognize one review's sentiment from two different perspectives. We experiment both W-SDDP and P-SDDP on 6 datasets, and compare the results with 6 other models. The results show that both W-SDDP and P-SDDP outperform other models, and this indicates that SDDP is a promising model for use in sentiment analysis.

2 Related Work

Aspect-based sentiment analysis has been deeply explored in the past decades, and most of them exploit Latent Dirichlet Allocation (LDA). Based on LDA, the word model and the phrase model are two research branches that have been drived[19].

Phrase Model processes the raw text data into a series of phrases, which can be shown as $< head\, word, modifier\, word >$. Generally speaking, the head word is used for identifying aspects, and the modifier word is for identifying sentiments. Several tries[12],[20] have been done, but because of the laborious data pre-process work, most researchers prefer the word model.

Word Model relies on the "bag-of-word" assumption. Two methods are commonly used to deal with words. First, some researches assume one word simultaneously conveys both sentiment and aspect information, so they use different aspect-sentiment prior distributions to infer the word assignment. Models like JST[8] and ASUM[21] belong to this category, which is referred as *Pure Word Model* in this paper. Second, external knowledge, like POS tagging or sentiment dictionary, might be used to help distinguish whether a word conveys aspect or sentiment information, and use aspect words to infer the aspects, and sentiment words to infer the sentiments. Models like JAS[22] and MaxEnt-LDA[10] fall into this category. Comparing to the former category, these models is closer to phrase model, since it also needs much data pre-process work. The only difference is that it does not need to pair up the aspect and sentiment words. This paper refers this kind of word model as *Mixture Word Model*.

Currently, most models are LDA based. The problem of LDA extended models is that the number of aspects needs to be pre-determined. However, such a decision is hard to make, especially in the social media environment. For example, one can hardly know how many aspects people may mention about towards a restaurant in Yelp.com. Traditionally, researchers will take a lot of trials and errors to find the optimal number, but it is time and labor consuming. One solution to this problem is to replace the Dirichlet allocation in LDA with Dirichlet process, which can help to generate the number of aspects automatically. Thus, Hierarchical Dirichlet Process(HDP)[17] was introduced as a non-parametric model to solve the aspect number determination problem.

HDP consists of two levels of Dirichlet Process(DP). The second level is a multinomial distribution, which is represented as G_0, and G_0 is shared by all the documents in the data collection. Specifically,G_0 can be deemed as the aspect distributions we want to generate from the review collection. The first level is a series of multinomial distributions θ_d, which control word-aspect assignments within each document. θ_d are private to each of the documents, and generated by G_0. If we explain HDP from the Chinese Restaurant Process(CRP)[23] perspective, each document can be treated as a restaurant, and words in the document are the customers, who will sit tables within this restaurant. The first customer will sit the first table, and the n^{th} customer will sit at table t with probability of $\frac{c_t}{n-1+\alpha}$, where c_t is the number of words sat in table t, and create a new table with probability of $\frac{\alpha}{n-1+\alpha}$. Similarly, tables will order dishes. A table will order a dish k with probability of $\frac{s_k}{m-1+\gamma}$, where s_k is the number of tables ordered dish k, and create a new dish with probability of $\frac{\gamma}{m-1+\gamma}$. α and γ are the hyper-parameters in the model. The table and dish here can be treated as the local and global aspect distributions. The graphic model of HDP is shown in Figure 1(a).

Just as shown in the model, the probability of a customer sitting at a table is only proportional to the number of other customers already in that table[24]. Such assignments are kind of random and ignore the context information. Aiming to solve this problem, Blei has proposed a Distance Dependency Chinese Restaurant Process(Dist-CRP) [24], which takes the distance information into consideration in image process. Although it improves the performance by clustering close pixels within one document, Dist-CRP ignores the words' global co-occurrence information. Actually, one reason why LDA can achieve a satisfying result in topic detection is that it exploits the words' co-occurrence information well, while HDP, including Dist-CRP, has overlooked such information by implementing the private $\boldsymbol{\theta_d}s$. Thus, this paper has proposed a Semantic Dependency Dirichlet Process (SDDP) for sentiment analysis. SDDP considers not only the local distance information, but also the global co-occurrence information. Based on SDDP, we construct two kinds of models, one word model (W-SDDP) and one phrase model (P-SDDP). The experiment results show that SDDP based models perform better than LDA, HDP and their extended models.

3 Model Description.

3.1 Similarity Dependency Dirichlet Process.

Words' co-occurrence and distance are the two perspectives reflecting two words' semantic similarity. Thus, this paper uses formula(3.1) as the function to calculate two words' semantic similarity.

(3.1)

$$sim(w_i, w_j) = m * \sum_{d=1}^{D} \sum_{i,j=0}^{M_d} \left(\frac{e^{-|i-j|} - e^{-M_d}}{e^{-1} - e^{-M_d}} * \frac{1}{c(w_i) + c(w_j)}\right)$$

where D is the number of Document, M_d is the length of document d, w_i represents the word appearing in document d's i^{th} position, $c(w_i)$ denotes the total number of times w_i occurring in the document collection, and m is the normalization coefficient to ensure $sim(w_i, w_j)$ ranges in [0,1]. Thus, the more often w_i and w_j appear in the same document, or the closer w_i and w_j present, the larger $sim(w_i, w_j)$ is.

SDDP utilize $sim(w_i, w_j)$ to assign words. Given $w_1, w_2, .w_{n-1}$ within a document, the table assignment a_n of the n^{th} word/phrase follows the formula (3.2), where t represents the table assignments, and $count(t)$ means the number of words that have been assigned to table t.

$$p(a_n = t|\boldsymbol{a_{1:n-1}}, \boldsymbol{w_{i\in t}}, \alpha, sim(\cdot)) \propto \tag{3.2}$$
$$\begin{cases} \dfrac{\sum_{i\in t} sim(w_i, w_n)}{count(t)} & \text{(if t exists)} \\ \alpha & \text{(if t is new)} \end{cases}$$

Similarly, in the second level, given the all the tables $t_1, t_2, .t_{m-1}$ generated among the whole document collection, the topic assignment z_m for the m^{th} table can be represented as formula(3.3), where k represents the topic assignment, and $count(k)$ is the number of tables assigned to this topic.

$$p(z_m = k|\boldsymbol{z_{1:m-1}}, \boldsymbol{t_{j\in k}}, \gamma, sim(\cdot)) \propto \tag{3.3}$$
$$\begin{cases} \dfrac{\sum_{j\in k} sim(t_j, t_m)}{count(k)} & \text{(if k exists)} \\ \gamma & \text{(if k is new)} \end{cases}$$

3.2 Two Derivation Models. Just as mentioned in related work part, there are mainly two types of models dealing with sentiment analysis, namely the word model and the phrase model, and in addition, the word model

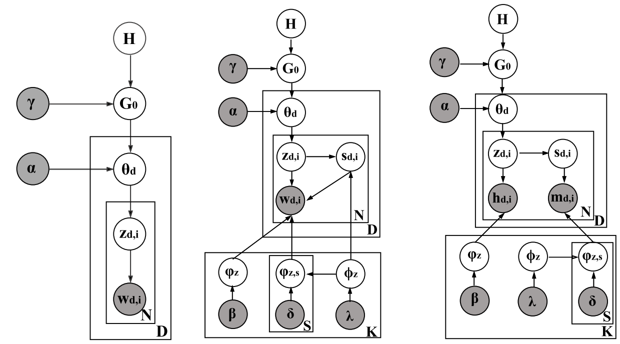


Figure 1: Graphic Models

can be classified into the pure word model and the mixture word model. Considering the mixture word model is a kind of combination of the phrase model and the pure word model, and in order to make a complete contrast, this paper extends SDDP into a pure word model (W-SDDP) and a phrase model (P-SDDP).

W-SDDP assumes each document is composed by a series of independent words, and each word conveys both aspect and sentiment information. While, P-SDDP assumes each document is composed by a series of independent phrase, which can be represented as $< head\ word, modifier\ word >$, and the head word conveys the aspect information, and the modifier word conveys the sentiment information. Figure 1(b) and Figure 1(c) show the graphic models for W-SDDP and P-SDDP. Table 1 shows the explanations for the annotation in Figure 1.

In generative process, W-SDDP and P-SDDP have a lot in common. To be concise, we use the same flow to introduce the generative process, but highlight the different parts for these two models. Their generative process can be shown as follows:

Step 1: Define a baseline H for global aspect generation. Here we choose a uniform distribution as H. Draw a distribution G_0 from H according to SDDP parametrized by γ. $G_0 \sim SDDP(H, \gamma)$.

Step 2: For each aspect, draw a word-aspect distribution φ_k according to a Dirichlet distribution parametrized by β. $\varphi_k \sim Dir(\beta)$.

Step 3: For each aspect, draw sentiment distributions $\varphi_{k,s}$ according to a Dirichlet distribution parametrized by δ_s. $\varphi_{k,s} \sim Dir(\delta_s)$.

Step 4: For each document d:

(4.1) Draw a multinomial distribution θ_d from G_0 according to SDDP parametrized by α. $\theta_d \sim SDDP(G_0, \alpha)$.

(4.2) For the i^{th} word or phrase in document d:

(4.2.1) Draw an aspect assignment $z_{d,i}$ according to θ_d.

(4.2.2) Draw a sentiment distribution ϕ_z according to a Dirichlet distribution parametrized by λ. $\phi_z \sim Dir(\lambda)$.

(4.2.3) Draw a sentiment assignment $s_{d,i}$ according to ϕ_z.

Table 1: Annotation of the Graphic Model

D	The number of documents
N	The number of words/phrases
K	The number of aspects
S	The number of sentiments
H	The baseline distribution to generate G_0
G_0	The global aspect distribution shared by all the documents
θ_d	The The local aspect distribution of document d
φ_z	The word distribution within an aspect z
$\varphi_{z,s}$	The word distribution of a sentiment s in aspect z
ϕ_z	The sentiment distribution within an aspect z
$z_{d,i}$	The aspect assignment of the i^{th} word/phrase in the d^{th} document
$s_{d,i}$	The sentiment assignment of the i^{th} word/phrase in the d^{th} document
$w_{d,i}$	The i^{th} word in the d^{th} document
$h_{d,i}$	The head word of the i^{th} phrase in the d^{th} document
$m_{d,i}$	The modifier word of the i^{th} phrase in the d^{th} document
α	The hyper parameter for local aspect assignment
β	The hyper parameter for word allocation within an aspect
γ	The hyper parameter for global aspect assignment
λ	The hyper parameter for sentiment distribution within an aspect
δ	The hyper parameter for word distribution within a sentiment

For W-SDDP:

(4.2.4) Generate a word $w_{d,i}$ according to φ_z and $\varphi_{z,s}$. $w_{d,i} \sim \varphi_z, \varphi_{z,s}$.

For P-SDDP:

(4.2.4) Generate the head of $p_{d,i}$ according to φ_z. $h_{d,i} \sim \varphi_z$.

(4.2.5) Generate the modifier of $p_{d,i}$ according to $\varphi_{z,s}$. $m_{d,i} \sim \varphi_{z,s}$.

3.3 Model Inference. We use Gibbs Sampling to infer both of W-SDDP and P-SDDP. Considering W-SDDP and P-SDDP share very similar sampling process, and the only difference is that W-SDDP use words to sample both aspects and sentiments, while P-SDDP use head words to infer aspects and modifier words to infer sentiments. To make the paper concise, we use W-SDDP as the example to show how to make the inference, as for P-SDDP, please replace $w_{d,i}$ with $h_{d,i}$ in aspect inference, and replace $w_{d,i}$ with $m_{d,i}$ in sentiment inference.

At the first level, each word needs to be assigned to different tables. This assignment can be realized by the formula(3.4), where $a_{d,i}$ denotes the table assignment of $w_{d,i}$, $\boldsymbol{a^{-d,i}}$ represents other words' table assignments in document d except $w_{d,i}$, k represents the aspects' word distributions, $ave(sim(w_{d,i}, w_n))$ denotes to $\frac{\sum_{i\in t} sim(w_n,w_i)}{count(t)}$in formula(3.2), and $g_k^{-w_{d,i}}(k_t)$ denotes to the word distribution in the aspect which table t has been assigned to.

$$p(a_{d,i} = t|\boldsymbol{a^{-d,i}}, \boldsymbol{k}) =$$

$$(\sum_{t=1}^{T} \frac{\sum_{n\in t} ave(sim(w_{d,i}, w_n))}{\sum_{t=1}^{T}\sum_{n\in t} ave(sim(w_{d,i}, w_n)) + \alpha} * g_k^{-w_{d,i}}(k_t)) + \frac{\alpha}{\sum_{t=1}^{T}\sum_{n\in t} ave(sim(w_{d,i}, w_n)) + \alpha} \tag{3.4}$$

Similarly, table will be assigned to different aspects in the second level, the inference process can be shown as formula (3.5), where $z_{d,t}$ denotes to the aspect assignment for the t^{th} table in document d, $ave(sim(t_{d,j}, t_m))$ denotes to $\frac{\sum_{m\in k} sim(t_m,t_j)}{count(k)}$ in formula (3.3), and $g_k^{-t_{d,j}}(k)$ denotes to the word distribution in aspect k.

$$p(z_{d,j} = k|\boldsymbol{z^{-d,j}}, \boldsymbol{k}) =$$

$$(\sum_{k=1}^{K} \frac{\sum_{m\in k} ave(sim(t_{d,j}, t_m))}{\sum_{k=1}^{K}\sum_{m\in k} ave(sim(t_{d,j}, t_m)) + \gamma} * g_k^{-t_{d,j}}(k)) + \frac{\gamma}{\sum_{k=1}^{K}\sum_{t\in k} ave(sim(t_{d,j}, t_m)) + \gamma} \tag{3.5}$$

$g_k^{-w_{d,i}}(k)$can be inferred as formula(3.6) , where $N_{k,w}^{-w_{d,i}}$ is the word count of w in aspect k except $w_{d,i}$, $N_k^{-w_{d,i}}$ is the number of words in aspect k except $w_{d,i}$ and V is the length of word vocabulary.

$$g_k^{-w_{d,i}}(k) = \frac{N_{k,w}^{-w_{d,i}} + \beta}{N_k^{-w_{d,i}} + V * \beta} \tag{3.6}$$

After aspect inference, a sentiment needs to be chosen for the very word or phrase under this aspect. We apply a Dirichlet allocation as the prior for sentiment distribution. Under the aspect k, the sentiment inference can be made as formula (3.7), where $N_{k,s,w}^{-w_{d,i}}$is the number of word w has been assigned to sentiment s under aspect k except $w_{d,i}$, $N_{k,s}^{-w_{d,i}}$ the number of words have been assigned to sentiment s under aspect k except $w_{d,i}$, $N_k^{-w_{d,i}}$ is the word count of aspect k except $w_{d,i}$, and S is the number of sentiment.

$$p(s_{d,i} = s|k) = \frac{N_{k,s,w}^{-w_{d,i}} + \delta_s}{N_{k,s}^{-w_{d,i}} + V * S} * \frac{N_{k,s}^{-w_{d,i}} + \lambda}{N_k^{-w_{d,i}} + S * \lambda} \tag{3.7}$$

So, in summary the inference process can be represented as the formula (3.8).

$$p(z_{d,i} = k, s_{d,i} = s|\boldsymbol{parameters}) =$$

$$\begin{aligned} &((\sum_{t=1}^{T} \frac{\sum_{n\in t} ave(sim(w_{d,i}, w_n))}{\sum_{t=1}^{T}\sum_{n\in t} ave(sim(w_{d,j}, w_n)) + \alpha} * g_k^{-w_{d,i}}(k_t)) \\ &+ \frac{\alpha}{\sum_{t=1}^{T}\sum_{n\in t} ave(sim(w_{d,i}, w_n)) + \alpha} \\ &* ((\sum_{k=1}^{K} \frac{\sum_{m\in k} ave(sim(t_{d,i}, t_m))}{\sum_{k=1}^{K}\sum_{m\in k} ave(sim(t_{d,i}, t_m)) + \gamma} * g_k^{-t_{d,i}}(k)) \\ &+ \frac{\gamma}{\sum_{k=1}^{K}\sum_{m\in k} ave(sim(t_{d,i}, t_m)) + \gamma} * G_0)) \\ &* (\frac{N_{k,s,w}^{-w_{d,i}} + \delta_s}{N_{k,s}^{-w_{d,i}} + V * S} * \frac{N_{k,s}^{-w_{d,i}} + \lambda}{N_k^{-w_{d,i}} + S * \lambda}) \end{aligned} \tag{3.8}$$

4 Experiment and Evaluation.

4.1 Dataset and Model Description. All the experiments and evaluations are tested on one or some of the six datasets shown in Table 2. We used different datasets for different experiment or evaluation purposes. The models to be tested are LDA[7], HDP[17], JST[8], ASUM[11], MaxEnt-LDA[10], JAS[22], and our two models: W-SDDP and P-SDDP.

Table 2: Dataset List

NO.	Dataset Content	Source	Volume	Labelled
1[25]	Restaurant	Citysearch	3400 sentences	Yes
2[21]	Coffee Machine	Amazon	3000 reviews	No
3[21]	Laptop	Amazon	3000 reviews	No
4[26]	Car	tripAdviser	3000 reviews	No
5[26]	Hotel	tripAdviser	3000 reviews	No
6	Restaurant	Yelp.com	350 reviews	No

4.2 Experiment Settings. P-SDDP functions only when the documents can be decomposed into a series of phrases, which are presented as $< head\,word, modifier\,word >$. We use Stanford Dependency Parser (SDParser)[27] to process the datasets for phrase model. Given a sentence, SDParser can find the word pairs that have modification relations. According to the results provided by SDParser, the overall relationships and patterns we use are listed as follows, where A in $< A, B >$ denotes the head word and B in $< A, B >$ denotes the modifier word. For detailed information, please refer to SDParser manual book[28].

$AdjectivalModifier : amod(A, B) \rightarrow < A, B >$

$Adjectival \quad Complement : acomp(A, B) + nsubj(A, C) \rightarrow < C, B >$

$Copula : cop(A, B) + nsubj(A, C) \rightarrow < C, A >$

$Direct \quad Object : dobj(A, B) + nsubj(A, C) \rightarrow < B, A >$

$And :< A, B > +conj\_and(A, C) \rightarrow < C, B > or < A, B > +conj\_and(B, C) \rightarrow < A, C >$

$Negation \quad Modifier :< A, B > +neg(B, not) \rightarrow < A, not + B >$

$Noun \quad Compound :< A, B > +nn(A, C) \rightarrow < C + A, B >, or < A, B > +nn < C, A \rightarrow < A + C, B >$

$Agent \quad Relationship : agent(A, B) \rightarrow < B, A >$

$Nominal \quad Subject : nsubj(A, B) \rightarrow < B, A >$

$Infinitival \quad Modifier : infmod(A, B) \rightarrow < A, B >$

$Passive \quad Nominal \quad Subject : nsubjpass < A, B > \rightarrow < B, A >$

$Participial \quad Modifier : partmod(A, B) \rightarrow < A, B >$

$Controlling \quad Subject : xsubj(A, B) \rightarrow < B, A >$

4.3 Prior Knowledge. We use MQPA as the sentiment dictionary to facilitate the hyper parameter settings for sentiment distribution. MQPA has provided the polarities for each word, but not the explicit score. Thus, we make the following rules:

If a word is tagged as "positive" and "strongsubj", $\delta_{positive} = 0.8, \delta_{negative} = 0.1, and, \delta_{neutral} = 0.1$

If a word is tagged as "positive" and "weaksubj", $\delta_{positive} = 0.6, \delta_{negative} = 0.1, and, \delta_{neutral} = 0.3$

If a word is tagged as "negative" and "strongsubj", $\delta_{positive} = 0.1, \delta_{negative} = 0.8, and, \delta_{neutral} = 0.1$

If a word is tagged as "negative" and "weaksubj", $\delta_{positive} = 0.1, \delta_{negative} = 0.6, and, \delta_{neutral} = 0.3$

If a word is tagged as "neutral" and "strongsubj", $\delta_{positive} = 0.1, \delta_{negative} = 0.1, and, \delta_{neutral} = 0.8$

If a word is tagged as "neutral" and "weaksubj", $\lambda_{positive} = 0.6, \lambda_{negative} = 0.2, and, \lambda_{neutral} = 0.2$

For other hyper parameters in the models, we employ the standard and out of box settings without any tuning to our datasets. For the six comparison models, all the hyper parameters are set as default values, and for W-SDDP, and P-SDDP, α and γ are set with the same magnitude of the similarity value, and β=λ=0.05.

4.4 Evaluation

Evaluation with Golden Standard. First, we use golden standard to evaluate our models. The first dataset in Table 2 is the one with golden standard. All the words in this dataset have been manually annotated to six aspects, namely Food, Staff, Price, Ambience, Anecdote, and Miscellaneous, and three sentiments: Positive, Negative and Neutral.

Models like JST, ASUM, and JAS, mix all the aspect words and sentiment words together, so we extract the noun words as the aspect words and others as sentiment words to map with the golden standard. Models like MaxEnt, LDA and HDP, do not provide sentiment polarities, so we can only compare them on the general sentiment level without distinguishing the specific sentiment types.

We use precision to measure to what degree each model can correctly select the aspect words and the corresponding sentiment words. The results are shown in Table 3, Table 4 and Table 5. The blanks in the tables mean that we could not find the related aspect or sentiment from the model's results.

Table 3: Aspect Comparison among the Popular Models

	LDA	HDP	ASUM	JST	Max-Ent	JAS	W-SDDP	P-SDDP
Food	0.639	0.806	0.751	0.632	0.808	0.779	0.760	**0.817**
Staff	0.429	0.460	0.411	0.299	0.559	0.527	0.563	**0.655**
Price	–	0.353	0.278	–	0.232	0.351	0.366	**0.494**
Ambience	0.412	0.452	0.347	0.226	0.299	0.451	0.469	**0.545**
Anecdote	0.379	0.444	0.259	0.188	0.397	0.443	0.450	**0.450**
Miscellaneous	0.441	0.471	0.504	0.347	0.330	0.532	0.565	**0.590**

From Table 3 we can find that P-SDDP performs the best in identifying all the six aspects, and W-SDDP also outperforms other non-SDDP based models except in detecting "Food", which is just slightly lower.

Table 4: Sentiment Comparison among the Models with Sentiment Polarity

		ASUM	JST	JAS	W-SDDP	P-SDDP
Food	+	0.655	0.461	0.658	**0.822**	0.786
	-	0.368	0.225	0.224	**0.440**	0.400
	*	0.104	0.064	–	0.136	**0.304**
Staff	+	0.445	0.241	0.243	**0.667**	0.662
	-	0.388	0.164	0.322	0.438	**0.651**
	*	0.022	0.037	–	0.071	0.063
Price	+	–	–	0.255	0.333	**0.431**
	-	0.150	–	0.088	**0.333**	0.273
	*	–	–	–	0.000	0.000
Ambience	+	–	–	0.273	**0.701**	0.565
	-	0.174	–	0.124	0.286	**0.400**
	*	0.056	0.029	–	0.078	**0.158**
Anecdote	+	–	0.089	0.093	**0.500**	0.256
	-	–	–	0.143	**0.333**	0.250
	*	0.243	0.113	–	0.200	**0.444**
Miscellaneous	+	0.302	0.241	0.227	**0.636**	0.583
	-	0.218	–	0.176	0.250	**0.400**
	*	0.219	–	–	**0.500**	0.231

Table 5: Sentiment Comparison among Models without Sentiment Polarity

	LDA	HDP	Max-Ent	W-SDDP	P-SDDP
Food	0.230	0.161	0.221	**0.602**	0.530
Staff	0.197	0.090	0.205	**0.583**	0.391
Price	–	0.059	0.134	**0.301**	0.263
Ambience	0.187	0.082	0.107	**0.440**	0.406
Anecdote	0.164	0.083	0.131	0.281	**0.333**
Miscellaneous	0.190	0.000	0.091	0.452	**0.500**

From Table 4 and 5, we can find the both W-SDDP and P-SDDP outperform other models, and they beat each other alternatively.

In this section, we find that both of W-SDDP and P-SDDP perform better than other models when evaluating via datasets with golden standard.

Evaluation with Perplexity. Although we can evaluate models via golden standard, the quantity of labelled datasets is very limited. It is not persuasive by just testing the results on only one dataset. Another way to evaluate the performances of the models is to apply *Perplexity* on unlabelled datasets.

We test the 8 models on 4 datasets, from the 2^{nd} to 5^{th} dataset in Table 2. For each model on each dataset, we change the initial number of aspects from 10 to 100 in intervals of 10, and choose the lowest perplexity as the value to compare. The results are shown in Figure 2. From Figure 2, we can see that P-SDDP always has the lowest perplexity on all the datasets, followed by W-SDDP. Perplexity based evaluation indicates that

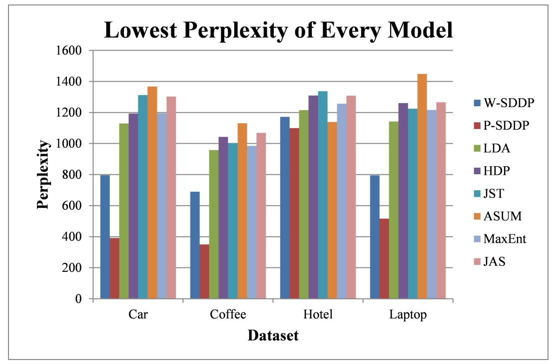


Figure 2: Perplexity Comparison among Models

W-SDDP and P-SDDP have better performances in inferring the words in the datasets. Comparing to W-SDDP, P-SDDP has an even better performance.

5 Applications in Recommendation System.

In this section, we will show the practical applications of W-SDDP and P-SDDP in Recommendation System.

5.1 Aspect Specific Dictionary Construction. Both W-SDDP and P-SDDP can help to construct the aspect- specific dictionary. Table 7 and Table 8 show the major aspects and their corresponding sentiment words detected by W-SDDP and P-SDDP respectively from the first dataset in Table 2. From Table 7 and Table 8, we find for a restaurant, the commonly mentioned aspects are Food (including Chinese food, Japanese food etc.), Atmosphere, Service and Staff. In addition, for each aspect, people might use different sentiment words to express their feelings. The sentiment word like "Oily" conveys a negative sentiment for Chinese food, but "watery" conveys a positive sentiment. Similarly, the word "Repeatable" conveys a negative sentiment when describes the staff, but it may convey a positive sentiment in other scenario.This result can be implemented in Review Recommendation System to help detect the good reviews to recommend to customers.

5.2 Online Review Summarization. In many cases, we want to be recommend a product by a certain aspect. Taking restaurant as an example, by explicit providing the aspect and sentiment summarization information can help to recommend a restaurant more precisely.

We implement both W-SDDP and P-SDDP on the

Table 6: Result of W-SDDP

Aspect		Sentiment
Atmosphere-Service: Service,Place,Time, Menu,Atmosphere, Staff,Dishes,Drinks	+	Nice,Great,Wonderful,Decent, Popular,Relax,Superb,Friendly
	-	Dim,Horrible,Mediocre Disappointing,Crowded,Poorly Slow,Worst
Food-Pizza: Pizza,Crust,Slice, Cheese,Williamsburg, Mushroom	+	Adorable,Delicate,Crisp,Fancy Best,Pretty,Supreme,Perfect
	-	Horrific,Vomit,Disgusting Complaints,Tiny,Gross, Expensive, Not-Special
Food-Japan& China: Sushi,Sichuan,Roll Eel, Sea, Chongqing, Fish, Chinatown Shanghai	+	Heavenly,Rejoice,Special,Best, Amazingly,Favorite,Fresh, Elegant
	-	Mock, Rigid, Dull, Overdone, Fatty, Weird, Poor, Not-Fresh
Food-USA: Bagel, Bagels, Coffee, Freeze, Cream Cheeses,Takeaway Mayo	+	Colossal, Outstanding, Best, Plentiful, Big, Original, Pleasantly, Fabulous
	-	Strange, Pricey, Not-Nice, Not-Authentic, Bland, Spot, Disappointed
Staff: Table, Dinner, Waitstaff, Minute, Service, Minutes, Bartender, Waiter	+	Hospitable, Experienced, Nice, Stylish, Not-Unable, Helpful, Ready, Attentive
	-	Confused, Not-Amazed, Annoying, Not-Competent, Unpleasant, Noisy, Clumsy, Pretentious

Table 7: Result of P-SDDP

Aspect		Sentiment
Atmosphere-Service: Service, Place, Dishes, Atmosphere, Night, Staff	+	Reasonable, Accommodating, Friendly, Relaxing, Romantic, Excellent, Expected, Cool
	-	Rude, Noisy, Disappointing, Biting, Dark, Poor, Drafty, Slow
Food-Pizza: Pizza,Slice,Crust, Ingredients,Codfish Addition,Lobster,Pie	+	Crisp, Fresh, Thin, Expanded, Fresh-Tasting,Well-Seasoned, Delicious,Tasty
	-	Shredded, Vomit-Inducting, Not-Topped,Skimp,Not-Want, Common,Bitter,Bland
Food-Japan: Sushi,Rice,Tuna, Fish, Sauces, Scallop, Roll,Appetizer	+	Spicy,Matches,Please, Healthy-Looking, Recommended, Favorite Refreshing, Superb
	-	Disgusting,Flavorless,Not-Exciting,Broken,Horrid, Rough,Murky,Awful
Food-China: Pork,Soup, Dumpling,Chicken, Shanghai, Shanghainese, Scallion, Eggplant	+	Tasting,Traditional,Amazing, Watery,Love,Wonderful, Authentic,Complimentary
	-	Sour, Mock, Lacking, Horrible, Overcompensate, Oily, Overpriced,Small
Staff: Staff, Service, Manager,People, Cooks,Menu,Tables, Reservation	+	Friendly, Great, Enthusiastic, Attentive, Helpful, Knowledgeable, Wonderful
	-	Not-recommend,Lies,Bad, Unavailable, Repeatable, Unpleasant, Not-inspired, Lazy

6^{th} dataset, and find there are 5 aspects people mention a lot toward this restaurant, namely Chicken & Waffles, the signature dish of this restaurant, Food rather than Chicken & Waffles, Atmosphere, Service and Price. For most aspects, like food (including Chicken & Waffles), atmosphere, and service, people tend to give a positive judgement. While for the price, the negative sentiment proportion is a little larger. Thus, to a consumer who emphasize the food or service quality, we can recommend this restaurant, but to a consumer who cares about the price, we may ignore this restaurant.

6 Discussion and Conclusion.

6.1 Comparison between W-SDDP and P-SDDP. Section 4.4 has proved that W-SDDP and P-SDDP indeed outperform other models. In this part, we will compare W-SDDP and P-SDDP to see which one is better in application. All the experiments in this part are conducted on the first dataset in Table 2.

Table 6 shows that comparing to W-SDDP, P-SDDP has a lower converged aspect number and a lower perplexity. In this sense, P-SDDP performs better than W-SDDP.However, P-SDDP loses a lot of information. In Table 6, we can see that P-SDDP loses near to 32.5% word token in phrase process, because some words could not be paired up to phrases, and are removed by the parser.

Thus, in real use, one needs to balance these two models. P-SDDP will give a more concise and better result, but lose considerable amount of information. W-SDDP keeps all the information, but might bring some noise to the results.

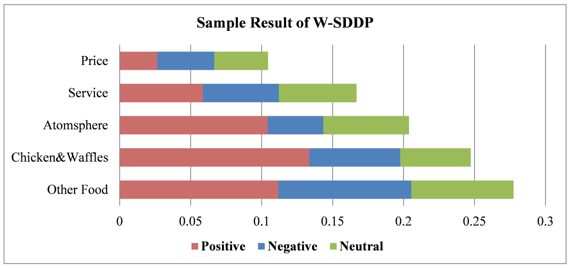
Table 8: Comparison between W-SDDP and P-SDDP

	W-SDDP	P-SDDP
Number Of Tokens	30035	20274
Converged Aspect Number	20-30	8-10
Perplexity	Around 900	Around 300

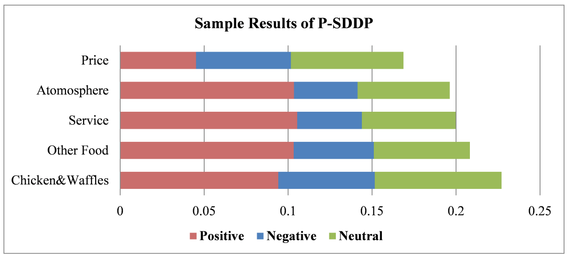
6.2 Conclusion. *Sentiment Analysis* is a core component for review recommendation system. This paper has constructed a Similarity Dependency Dirichlet Process (SDDP) as a novel model for sentiment analysis. SDDP has solved the aspect number specification problem encountered in LDA, and improves the aspect/sentiment detection performance by replacing the random word assignment mechanism with similarity based word assignment.Based on SDDP, two models are built. One is a word model W-SDDP, and a phrase model P-SDDP. Evaluation results show that both W-SDDP and P-SDDP perform well on various datasets.

References

[1] Ghose, A, P. G. Ipeirotis,*Designing novel review ranking systems: predicting the usefulness and impact of reviews,*



(a) Sample Result of W-SDDP



(b) Sample Result of P-SDDP

Figure 3: Sample Results

Proceedings of the ninth international conference on Electronic commerce, ACM, 2007.

[2] Momeni, E., K. Tao, B. Haslhofer,G.-J. Houben, *Identification of Useful User Comments in Social Media: A Case Study on Flickr Commons*,Joint Confernece on Digital Library, 2013.

[3] S. R. Das and M. Y. Chen,*Yahoo! for Amazon: Sentiment extraction from small talk on the web*, Management Science, vol. 53, pp. 1375-1388, 2007.

[4] L. Dini and G. Mazzini,*Opinion classification through information extraction*, in Proceedings of the Conference on Data Mining Methods and Databases for Engineering, Finance and Other Fields (Data Mining), 2002, pp. 299-310.

[5] T. Nasukawa and J. Yi,*Sentiment analysis: Capturing favorability using natural language processing*, in Proceedings of the 2nd international conference on Knowledge capture, 2003, pp. 70-77.

[6] B. Pang, L. Lee, and S. Vaithyanathan,*Thumbs up?: sentiment classification using machine learning techniques*, in Proceedings of the ACL-02 conference on Empirical methods in natural language processing-Volume 10, 2002, pp. 79-86.

[7] D. M. Blei, A. Y. Ng, and M. I. Jordan,*Latent Dirichlet allocation*, Journal of Machine Learning Research, vol. 3, pp. 993-1022, May 15 2003.

[8] C. Lin and Y. He,*Joint sentiment/topic model for sentiment analysis*, in Proceedings of the 18th ACM conference on Information and knowledge management, 2009, pp. 375-384.

[9] I. Titov and R. McDonald,*Modeling online reviews with multi-grain topic models*, in Proceedings of the 17th international conference on World Wide Web, 2008, pp. 111-120.

[10] W. 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, 2010, pp. 56-65.

[11] S. Brody and N. Elhadad,*An unsupervised aspect sentiment model for online reviews*, in Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics, 2010, pp. 804-812.

[12] T. J. Zhan and C. H. Li,*Semantic Dependent Word Pairs Generative Model for Fine-Grained Product Feature Mining*, PAKDD 2011, Shenzhen China, May 24-27,2011, pp. 460-475, 2011.

[13] W. Jin, H. H. Ho, and R. K. Srihari,*OpinionMiner: a novel machine learning system for web opinion mining and extraction*, in Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining, 2009, pp. 1195-1204.

[14] H. Guo, H. Zhu, Z. Guo, and Z. Su,*Domain customization for aspect-oriented opinion analysis with multi-level latent sentiment clues*, in Proceedings of the 20th ACM international conference on Information and knowledge management, 2011, pp. 2493-2496.

[15] J. Si, A. Mukherjee, B. Liu, Q. Li, H. Li, and X. Deng,*Exploiting Topic based Twitter Sentiment for Stock Prediction*, in ACL, 2013, pp. 24-29.

[16] S. Kim, J. Zhang, Z. Chen, A. H. Oh, and S. Liu,*A Hierarchical Aspect-Sentiment Model for Online Reviews*, in AAAI, 2013.

[17] Y. W. Teh, M. I. Jordan, M. J. Beal, and D. M. Blei,*Hierarchical Dirichlet Processes*,Journal of the American Statistical Association, vol. 101, pp. 1566-1581, 2006.

[18] D. M. Blei and P. I. Frazier,*Distance dependent Chinese restaurant processes*, The Journal of Machine Learning Research, vol. 12, pp. 2461-2488, 2011.

[19] S. Moghaddam and M. Ester,*On the design of LDA models for aspect-based opinion mining*, in Proceedings of the 21st ACM international conference on Information and knowledge management, 2012, pp. 803-812.

[20] S. Moghaddam and M. Ester,*ILDA: Interdependent LDA Model for Learning Latent Aspects and Their Ratings From Online Product Reviews*, SIGIR'11, pp. 665-674, 2011.

[21] Y. Jo and A. Oh,*Aspect and Sentiment Unification Model for Online Review Analysis*, WSDM'11 February 9-12, Hong Kong, China, 2011.

[22] X. Xu, S. Tan, Y. Liu, X. Cheng, and Z. Lin,*Towards Jointly Extracting Aspects and Aspect-Specific Sentiment Knowledge*, CIKM'12, October 29-November , 2012 Maui, HI, USA, 2012.

[23] D. Blei,*COS 597C: Bayesian nonparametrics*,Lecture Notes in Priceton University. http:// www.cs.princeton.edu/ courses/ archive/ fall07/ cos597C/ scribe/ 20070921.pdf, 2007.

[24] D. M. Blei and P. Frazier*Distance Dependent Chinese Restaurant Process*, Proceedings of the 27th International Conference on Machine Learning, Haifa, Isreal, 2010.

[25] G. Ganu, N. Elhadad, and A. Marian,*Beyond the Stars: Improving Rating Predictions using Revew Text Cotent*, Twelfth International Workshop on the Web and Databases, Providence, Rhode, Island, USA, 2009.

[26] K. Ganesan and C. Zhai,*Opinion-based entity ranking*, Information retrieval, vol. 15, pp. 116-150, 2012.

[27] M.C. De Marneffe, B. MacCartney, and C. D. Manning,*Generating typed dependency parses from phrase structure parses*, in Proceedings of LREC, 2006, pp. 449-454.

[28] M.C. De Marneffe and C. D. Manning,*Stanford typed dependencies manual*, URL http://nlp. stanford. edu/software/dependencies manual. pdf, 2008.