Prior-Based Dual Additive Latent Dirichlet Allocation for User-Item Connected Documents

Wei Zhang **Jianyong Wang**
Tsinghua National Laboratory for Information Science and Technology (TNList)
Department of Computer Science and Technology, Tsinghua University, Beijing, China
zwei11@mails.tsinghua.edu.cn, jianyong@tsinghua.edu.cn

Abstract

User-item connected documents, such as customer reviews for specific items in online shopping website and user tips in location-based social networks, have become more and more prevalent recently. Inferring the topic distributions of user-item connected documents is beneficial for many applications, including document classification and summarization of users and items. While many different topic models have been proposed for modeling multiple text, most of them cannot account for the dual role of user-item connected documents (each document is related to one user and one item simultaneously) in topic distribution generation process. In this paper, we propose a novel probabilistic topic model called Prior-based Dual Additive Latent Dirichlet Allocation (PDA-LDA). It addresses the dual role of each document by associating its Dirichlet prior for topic distribution with user and item topic factors, which leads to a document-level asymmetric Dirichlet prior. In the experiments, we evaluate PDA-LDA on several real datasets and the results demonstrate that our model is effective in comparison to several other models, including held-out perplexity on modeling text and document classification application.

1 Introduction

User-item connected documents, which are written by users for specific items, occur in many scenarios. For example, in online shopping websites (e.g., Amazon[1]), users prefer to write reviews to express their attitudes towards some products that they have bought. These reviews have become a major reference for candidate buyers to make decisions. Another example is that in location-based social networks (e.g., Foursquare[2]), users always propose tips for some point-of-interests (POIs), which are beneficial for other users to gain knowledge about the POIs they have not visited before. In summary, user-item connected documents own dual roles, with each document associated with one user and one item simultaneously. They not only reflect the core concerns of users, but also indicate the characteristics of items. Without specific explanation, we also use documents to denote user-item connected documents for simplicity in the rest of this paper.

Characterizing content of documents is an important problem and has many applications in different fields, e.g., natural language processing and information retrieval. Topic models [Hofmann, 1999; Blei *et al.*, 2003b] are elegant ways to solve this problem by assuming each document has a topic distribution and each topic is represented as a distribution over words. By this way, documents are summarized by their associated topic distributions in a high level. For user-item connected documents, not only the learned topic distributions are useful as usual, but if the topic factors of users and items can be differentiated, then they can be used for creating their self-introduction profiles. These profiles are beneficial for other users to gain knowledge about items and retailers to understand what their consumers care about. Further, current popular recommender systems can utilize them to make explainable recommendations through content matching.

While many topic models have been proposed in the last decade, almost all of them are designed for some specific tasks and do not emphasize the dual role phenomenon of user-item connected documents, let alone automatically infer topic factors of users and items simultaneously. For one user-item connected document, its associated topic distribution should be influenced by its corresponding user and item together, and for a user or an item, its topics consist in multiple documents related to it. Therefore, the main challenge is to connect user, item, and reviews in a unified topic model while ensure different combinations of user and item pair tend to generate distinct topic distributions. Besides, it is better for the topic model to simultaneously infer user and item topic factors. Existing approaches such as author-topic model [Rosen-Zvi *et al.*, 2004] can only capture user's topic distribution but ignore item's topics.

To address the above issues, we propose a novel generative probabilistic topic model called Prior-based Dual Additive Latent Dirichlet Allocation (PDA-LDA) to model the user-item connected documents. This model is somewhat inspired by [Wallach *et al.*, 2009] which points out that asymmetric Dirichlet priors over topic distributions can lead to additional benefit for topic models than symmetric priors. In PDA-LDA, we account for the dual role phenomenon by associating Dirichlet prior of each document with their corre-

[1] http://www.amazon.com/
[2] https://foursquare.com/

sponding user and item. More specifically, PDA-LDA first assumes each user or item is represented with a topic factor. When a component of a topic factor takes a larger value, it means the corresponding user or item concentrates more on that topic. Then an exponential function is utilized to combine a pair of user and topic factor for each document through its additive property of parameters to form a new Dirichlet prior and thus every component of the prior takes positive value. As a result, each document has a different Dirichlet prior. The comparison between priors of Latent Dirichlet Allocation (LDA) [Blei *et al.*, 2003b], asymmetric Latent Dirichlet Allocation (AS-LDA) [Wallach *et al.*, 2009], and PDA-LDA is shown in Table 1.

Table 1: Difference of Dirichlet hyper-priors ($\alpha, \alpha_d \in \mathbb{R}^{\mathrm{K}}$).

LDA	$\mathrm{Dir}(\alpha), \alpha_i = \alpha_j (i, j \in \{1, \ldots, K\})$
AS-LDA	$\mathrm{Dir}(\alpha), \alpha_i \neq \alpha_j (i, j \in \{1, \ldots, K\})$
PDA-LDA	$\mathrm{Dir}(\alpha_d) (d \in \{1, \ldots, \lvert D \rvert\})$

In short, LDA is assigned a corpus-level symmetric prior, AS-LDA has a corpus-level asymmetric prior, while PDA-LDA owns document-level asymmetric priors. The main advantages of PDA-LDA lie in three aspects. First of all, it models text better by accounting for the dual role in topic generations. Second, it can automatically summarize users and items by the learned topic factors. And last, it constructs document-level priors with pairs of user and item topic factors. This enables prediction on test documents since after model learning procedure, user and item topic factors are known for these documents.

Contributions. In all, the main contributions of this paper are summarized as follows:

- We are the first to address the dual role phenomenon for topic generations of topic models in user-item connected documents to the best of our knowledge.
- We propose a new generative model called PDA-LDA which accounts for the dual role phenomenon by connecting user and item topic factors to Dirichlet priors through exponential additive transformation.
- We evaluated the proposed model on several real data sets to verify the effectiveness of PDA-LDA. The experimental results demonstrate that our model not only achieves superior held-out perplexity on test data, but generates better topics for its good performance in document classification application as well.

In what follows, detailed model specification is introduced in Section 2. Then we provide experimental results and some analysis in Section 3. Related works are briefly discussed in Section 4. In the last, we draw a conclusion for this work..

2 Approach

This section will be divided into three parts. The first part is mainly about the description of PDA-LDA, including the generative process of the model. Then we derive the Gibbs EM based learning algorithm for the model. Finally, a concise introduction to prediction on test data is given. Before we proceed, the main mathematical symbols we will use later are shown for reference in Table 2.

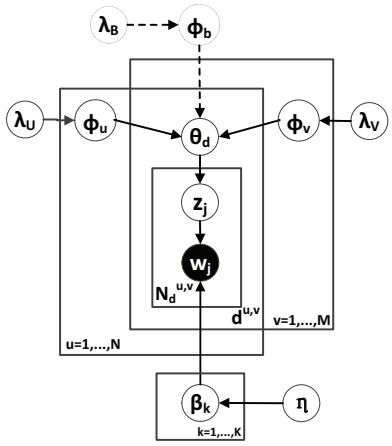


Figure 1: Graphical model of PDA-LDA.

2.1 Model Description

Standard topic models such as LDA always assume each document relates to a distribution over K latent components called topics and each topic can be explained by a distribution over all words in vocabulary W. Both the topic and word distributions are sampled from a Dirichlet distribution $\mathrm{Dir}(\alpha)$ known as prior in Bayesian learning. Our model follows the basic topic generation process of LDA. The key difference lies in how we set the parameter α of Dirichlet prior for each user-item connected document. The hyper-parameter α has been demonstrated to influence the result of text modeling [Wallach *et al.*, 2009].

As we discussed, each user or item has its own characteristics. We assume each user is associated with a topic factor $\phi_u \in \mathbb{R}^{\mathrm{K}}$ and the same for each item, i.e., $\phi_i \in \mathbb{R}^{\mathrm{K}}$. In topic factors, components with larger value denote users and items concentrate more on those corresponding topics. Topic factors should influence generation of topic distributions for documents. To bridge the gap between them, one natural idea is to associate topic factors with Dirichlet hyper-parameter α as in Bayesian learning framework, prior distribution reveals prior knowledge about how data will be generated before it really comes into action. This is consistent with the intrinsics of generation of user-item connected documents. Imagine that when a user plans to write something about an item, the first thing he will do is to choose some themes from what he often concentrates on and also conformed to the context of the item. Obviously, the themes are both relevant to the user and the item. These themes can be regarded as latent topics studied in topic models.

For a document $d^{u,v}$, we associate user and item topic factors with Dirichlet prior parameter of its topic distribution as follows,

$$\alpha_d^{u,v} = \exp(\phi_u + \phi_v) \tag{1}$$

where exponential transformation ensures each component of $\alpha_d^{u,v}$ to be positive, which satisfies the requirements of Dirichlet parameters. So far, topic factors of user and item can

Table 2: Notations of main adopted symbols.

D	User-item connected document set
W	Vocabulary set
$u \in U$	An user in the user set U
$v \in V$	An item in the item set V
K	Number of topics
λ_U	Precision of Gaussian prior for user factor ϕ_u
λ_V	Precision of Gaussian prior for item factor ϕ_v
λ_B	Precision of Gaussian prior for background factor ϕ_b
ϕ_u	Topic factor of user u
ϕ_v	Topic factor of item v
ϕ_b	Topic factor of background
α	Dirichlet prior parameter for topic distribution θ
θ_d	Topic distribution of document d
η	Dirichlet prior parameter for word distribution β
β_k	Multinomial distribution over words of topic k
z_j	Latent topic assignment for word w_j
w_j	The j-th word in document
$d^{u,v}$	The document written by user u for item v
$N_d^{u,v}$	Number of words in document $d^{u,v}$
$N_{d,k}^{u,v}$	Number of words with topic k in document
M_w^D	Occurrence of word w in corpus
$M_{w,k}^D$	Occurrence of word w with topic k in corpus

influence the generation of topic distribution of document $d^{u,v}$ through $\alpha_d^{u,v}$. One natural extension is to consider background topic factor ϕ_b additionally. It is necessary since some words such as conjunction words frequently occur in documents. The topic relates to these words are not specially owned by users or items. Therefore, the complete transformation between topic factors and Dirichlet parameter is defined to be

$$\alpha_d^{u,v} = \exp(\phi_u + \phi_v + \phi_b) \tag{2}$$

Based on the above analysis, we can summarize the generative story of PDA-LDA for user-item connected documents as below,

1. Draw background topic factor $\phi_b \sim \mathcal{N}(\lambda_B)$.
2. For each user u, draw user topic factor $\phi_u \sim \mathcal{N}(\lambda_U)$.
3. For each item v, draw item topic factor $\phi_v \sim \mathcal{N}(\lambda_V)$.
4. For each topic k, draw word distribution $\beta_k \sim$ Dirichlet(η).
5. For each user-item connected document $d^{u,v}$ in document collection D,
   (a) Draw topic distribution $\theta_d \sim$ Dirichlet$(\alpha_d^{u,v})$ where $\alpha_d^{u,v}$ is calculated through Equation(2).
   (b) For each position j in the document $d^{u,v}$:
   (b.1) Sample topic $z_j \sim$ Mult$(\theta_{d^{u,v}})$.
   (b.2) Draw word $w_j \sim$ Mult(β_{z_j}).

The complete graphical model representation of our proposed model is provided in Figure 1. The joint probability of a whole corpus D containing user-item connected documents and topic factors $\Theta_t = \{\phi_U, \phi_V, \phi_b\}$ is defined as follows,

$$\begin{aligned} P(D, \Theta_t) = \mathcal{N}(\phi_b; 0, \lambda_B) \prod_u \mathcal{N}(\phi_u; 0, \lambda_U) \prod_v \mathcal{N}(\phi_v; 0, \lambda_V) \\ \prod_{k'=1}^{K} \int Dir(\beta_{k'}; \eta) \prod_{d^{u,v}=1}^{D} \int Dir(\theta_d^{u,v}; \alpha_d^{u,v}) \\ \prod_{j \in N_d^{u,v}} \sum_{z_j} P(z_j | \theta_d^{u,v}) P(w_j | \beta_{z_j}) d\theta_d^{u,v} d\beta_{k'} \end{aligned} \tag{3}$$

2.2 Learning

In this work, our goal is to learn optimal topic factors Θ_t and key latent variables θ and β by maximizing the log-likelihood of joint probability shown in Equation (3),

$$\mathcal{L} = \log\big(P(D, \Theta_t)\big) \tag{4}$$

However, it is intractable to directly optimize the above function and compute the posterior distribution of θ and β due to the summarization of latent topics z_i in discrete space and coupling of θ and β. Luckily, if the assignments of latent topics can be inferred for all words in reviews, then not only θ and β can be calculated, but also Θ_t are tractable to be optimized. This intuition leads to the idea of Gibb EM algorithm.

Gibbs EM learning algorithm is widely adopted in (partially) Bayesian latent factor models [Wallach, 2006; Liu *et al.*, 2013] which alternates between sampling the value of latent variables (Expectation Step) and optimizing some model parameters (Maximization Step). More specifically, we adopt collapsed Gibbs Sampling to determine all the latent topics in E-step and optimize Θ_t through gradient ascent in M-step.

E-step

Given a word position j in document $d^{u,v}$, the key inferential problem in collapsed Gibbs sampling is to derive the posterior distribution of its associated topic, i.e., $P(z_{d,j}^{u,v} = k | \mathbf{z}_{d,-j}^{u,v}, \mathbf{w}_{\mathbf{d}}^{\mathbf{u},\mathbf{v}})$ where $\mathbf{z}_{d,-j}^{u,v}$ denotes all topic assignments of words in document d are known except the word in position j. As Θ_t are fixed in E-step, $\alpha_d^{u,v}$ can be calculated through Equation (2). By applying Bayesian formula and conditional dependence property, it is easy to derive the following formal definition of the posterior distribution according to the results of [Griffiths and Steyvers, 2004],

$$P(z_{d,j}^{u,v} = k | \mathbf{z}_{d,-j}^{u,v}, \mathbf{w}_{\mathbf{d}}^{\mathbf{u},\mathbf{v}}) = \frac{N_{d,k}^{u,v} + \alpha_{d,k}^{u,v}}{N_d + K\alpha_{d,k}^{u,v}} \frac{M_{w_j,k}^D + \eta}{M_w^D + |W|\eta} \tag{5}$$

where all the counts do not include the current word. The above equation is very similar to that in standard topic model except the Dirichlet prior $\alpha_{d,k}^{u,v}$.

M-step

After getting the latent topic assignments for all words in reviews, the objective log-likelihood of joint probability becomes $P(D, Z, \Theta_t)$ and thus the summarization term in Equation (3) vanishes. Due to the conjugate prior property of Dirichlet multinomial distribution [Heinrich, 2004; Griffiths and Steyvers, 2004], θ and β can be integrated out and $P(D, Z, \Theta_t)$ is now reformulated as below,

$$\begin{aligned} &P(D, Z, \Theta_t) = \\ &\mathcal{N}(\phi_b; 0, \lambda_B) \prod_u \mathcal{N}(\phi_u; 0, \lambda_U) \prod_v \mathcal{N}(\phi_v; 0, \lambda_V) \\ &\prod_{d^{u,v}} \frac{\prod_{k=1}^K \Gamma(N_{d,k}^{u,v} + \alpha_{d,k}^{u,v})}{\Gamma(N_d^{u,v} + \sum_{k=1}^K \alpha_{d,k}^{u,v})} \frac{\Gamma(\sum_{k=1}^K \alpha_{d,k}^{u,v})}{\prod_{k=1}^K \Gamma(\alpha_{d,k}^{u,v})} \\ &\prod_{k=1}^K \frac{\prod_w \Gamma(N_{d,k,w}^{u,v} + \eta)}{\Gamma(N_{d,k}^{u,v} + |W|\eta)} \frac{\Gamma(|W|\eta)}{\prod_w \Gamma(\eta)} \end{aligned} \tag{6}$$

where Γ is Gamma function with form $\Gamma(x) = (x-1)!$ when x is an integer.

By maximizing the log-likelihood of the above joint probability with gradient ascent algorithm, we can get optimal Θ_t. Gradient ascent consists of two steps. The first step is to educe the gradients of the parameters. For example, the gradient of $\phi_{u,k}$ is

$$\begin{aligned}\frac{\partial \mathcal{L}}{\partial \phi_{u,k}} = \sum_{u \in d^{u,v}} \alpha_{d,k}^{u,v} \Big(\Psi(\sum_{k'} \alpha_{d,k'}^{u,v}) - \Psi(\alpha_{d,k}^{u,v}) \\ + \Psi(N_{d,k}^{u,v} + \alpha_{d,k}^{u,v}) - \Psi(N_d^{u,v} + \sum_{k'} \alpha_{d,k'}^{u,v}) \Big) \\ - \lambda_U \phi_{u,k}\end{aligned} \tag{7}$$

where Ψ is Digamma function which is equal to logarithmic derivative of the Gamma function. The gradients of $\phi_{v,k}$ and $\phi_{b,k}$ are analogous to $\phi_{u,k}$ except the summarization condition and regularization term. We should emphasize although for one review, $\partial\phi_{u,k}$ and $\partial\phi_{v,k}$ are similar, different reviews connect to different users and items and thus the updates for the users and items are distinct from a whole perspective.

Based on the obtained gradients, the second step of gradient ascent is to update model parameters through

$$\Theta^{t+1} = \Theta^t + \omega \frac{\partial \mathcal{L}}{\partial \Theta^t} \tag{8}$$

where ω is the learning rate of the algorithm. Superscript t denotes the finished number of iterations.

Besides, after the iterative learning process converges, we can calculate β and θ by collecting enough subsequent samples though a burn-in process as [Blei *et al.*, 2003a],

$$\beta_{k,w} = \frac{\sum_t M_{t,w,k}^D + \eta}{\sum_t M_w^D + |W|\eta} \tag{9}$$

$$\theta_{d,k} = \frac{\sum_t N_{t,d,k}^{u,v} + \alpha_{d,k}^{u,v}}{\sum_t N_d + \sum_{k'} \alpha_{d,k'}^{u,v}} \tag{10}$$

where $N_{t,d,k}^{u,v}$ denotes the samples collected in the t iteration. In summary, the whole learning algorithm for PDA-LDA is concluded in Algorithm 1.

2.3 Prediction

When the PDA-LDA model is utilized for document modeling on test dataset, Θ_t and β should be fixed to be the optimal values learned from training set. The only important step is to infer latent topic assignments over test documents by sampling from a resembling formula as Equation (5), which additionally incorporates relevant count variables of test documents.

3 Experiments

In this section, we first introduce the datasets we used and the preprocessing steps for them. Then we discuss the adopted comparison models and the hyper-parameter setting of PDA-LDA. Finally, we analyze the experimental results to demonstrate the effectiveness of our method.

Algorithm 1: The Gibbs EM Algorithm for PDA-LDA

Input: User-item connect document collection D.
Output: Optimal user, item, and background topic factors Θ_t, document-topic distribution θ, topic-word distribution β

```
Initialize hyper-parameters and Θ_t
Randomly draw latent topic assignments of words in documents
tt = 0
while Not converged and tt ≤ Iter_max do
    E-step:
        (a) Sampling all words' latent topic assignments
            through Equation (5)
    M-step:
        (a) Calculate the gradients through Equation (7)
        (b) Update Θ_t through Equation (8)
    tt ← tt + 1
Calculate θ and β through Equation (9) and (10) under a burn-in process
```

3.1 Dataset

We adopt three real data collections from Yelp[3] and [McAuley and Leskovec, 2013]. Based on their origins, we denominate the three data sets as Yelp, AmazonFood and AmazonSport, respectively. To clean text data, we adopt the following four processing steps: (1) converting all letters into lowercase, (2) filtering stop words[4] and punctuations, (3) removing reviews which are too short, and (4) saving frequent words to form a vocabulary. After cleaning text data, we further remove users and items with less than 5 reviews to ensure that users and items have enough related documents. Finally, we obtain the experimental data sets whose basic statistics are shown in Table 3. We randomly divide the two collections into train, validation, and test set with the ratio $7:1:2$ for testing held-out perplexity and further binary document classification task.

Table 3: Introduction of Experimental Datasets.

Data	User	Item	Reviews	Length
Yelp	8017	5175	182139	52 (words)
AmazonFood	3681	1210	46053	63 (words)
AmazonSport	426	600	7982	53 (words)

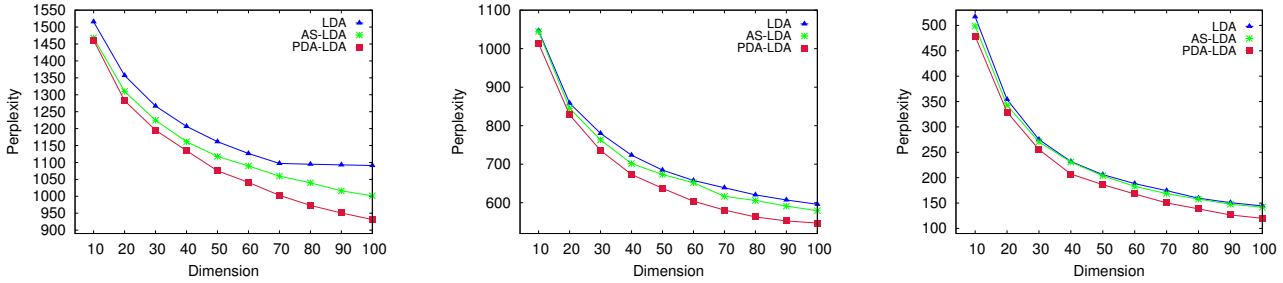
3.2 Comparison Models

To verify the effectiveness of our model, we first introduce the comparison models adopted in the experiments.

Latent Dirichlet Allocation (LDA): Although LDA [Blei *et al.*, 2003b] can neither capture author topics, nor obtain item topics, it is still necessary to analyze its result to verify the effectiveness of AS-LDA and PDA-LDA for exploring asymmetric Dirichlet prior as we mentioned in Section 1.

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

[4] http://jmlr.csail.mit.edu/papers/volume5/lewis04a/a11-smart-stop-list/english.stop



(a) Changed Perplexity on Yelp. (b) Changed Perplexity on AmazonFood. (c) Changed Perplexity on AmazonSport.

Figure 2: Result in terms of Perplexity with change of topic components.

Author-topic Model (ATM): We adapt ATM [Rosen-Zvi *et al.*, 2004] to our problem by regarding users as authors in the model. ATM somewhat resembles PDA-LDA as both of them can obtain users' topics. However, ATM cannot handle the double role phenomenon in user-item connected documents. Intuitively, it will perform worse than the new model.

Gibbs Sampler for Dirichlet Multinomial (GSDMM): GSDMM [Yin and Wang, 2014] is a standard topic model for handling short text such as tweets in social media by assuming all words in a document have the same topic and used as a comparison here due to the short length of the documents shown in Table 3,

Asymmetric Latent Dirichlet Allocation (AS-LDA): AS-LDA [Wallach *et al.*, 2009] is a strong comparison as it addresses popularity bias for each topic by incorporating asymmetric Dirichlet prior. It can be regarded as the most similar method to PDA-LDA which only considers the background topic factor for topic generation process.

3.3 Parameter Setting

All the hyper-parameters are determined based on their performances on validation datasets. For all the comparison methods and PDA-LDA, we assign 0.1 to η. For comparisons except AS-LDA, we choose α to be 0.1 as well for its good performance. The concentration parameter α' in AS-LDA is tunned to be 0.1. Apart from η, λ_U, λ_V, and λ_b are set to be 1 for PDA-LDA uniformly.

3.4 Results

Results on Text Modeling

We first compare all the adopted models in terms of perplexity on test datasets. Perplexity is widely used in probabilistic model for checking their quality. It is directly relevant to log-likelihood of probability and normally defined as below,

$$\texttt{Perplexity}(D_{test}) = \exp\left(-\frac{\sum_{d\in D_{test}} \log P(d)}{\sum_{d\in D_{test}} N_d}\right) \quad (11)$$

The detailed results are shown in Table 4. We notice that GSDMM performs not very well in user-item connected document modeling although the average length of the documents is short. One reason is that user-item connected documents are still longer than tweets in social media and very short text is more likely to contain only one topic.

Table 4: Comparisons of different models in terms of Perplexity on test data with K = 40.

Method	Yelp	AmazonFood	AmazonSport
LDA	1206.5	723.5	231.8
ATM	1788.7	1111.0	334.8
GSDMM	1521.7	1038.2	237.6
AS-LDA	1161.4	701.6	230.8
PDA-LDA	1134.7	673.2	207.2

ATM dose not show promising results although it models the role of users in topic generation. This may be explained by the fact that ATM is originally designed for the scenario where each document has multiple authors (e.g., research papers). As each user-item connected document is uniquely owned by one user, the ATM may not handle it very well.

AS-LDA outperforms LDA consistently in three datasets. This makes sense that asymmetric Dirichlet prior captures popularity bias of topics in generation process from a corpus level while LDA ignores the bias. It demonstrates that the idea of modifying hyper-parameters of Dirichlet prior for topic distribution is promising and also provides evidence for PDA-LDA to further differentiate prior parameters.

It is clear that PDA-LDA achieves superior performances among all the adopted models. Especially, PDA-LDA behaves better than the strong comparison AS-LDA significantly. This indicates that asymmetric Dirichlet prior which capture topic concentration bias from more grained document level is better than from corpus level. Moreover, we discover AS-LDA gains a minor improvement over LDA in the last dataset, while the decrease of perplexity of PDA-LDA is prominent, which indicates our model has wider applicability.

Dimension Influence on Text Modeling

We analyze the performance evolution with change of number of latent components from 10 to 100 with a step size of 10. We only compare LDA, AS-LDA, and PDA-LDA as their results are closer than the other two models. As Figure 2 shown, the perplexity of all the three models decrease steadily with increase of topic components. We choose $K = 40$ for other experiments in this work by considering a trade-off between efficiency and effectiveness as larger number of components costs more time to learn models. Besides, the perplexity dif-

Table 5: Comparisons of different methods in terms of Accuracy, Precision, and F1 Metric on classification task with K=40.

Method	Yelp			AmazonFood			AmazonSport		
	Accuracy	Precision	F1	Accuracy	Precision	F1	Accuracy	Precison	F1
LDA	0.770	0.765	0.748	0.729	0.692	0.653	0.768	0.752	0.748
AS-LDA	0.688	0.693	0.570	0.728	0.686	0.662	0.776	0.766	0.750
PDA-LDA	0.773	0.767	0.754	0.735	0.702	0.676	0.796	0.787	0.782

ferences between these models are relatively evident when $K \geq 40$ and thus the comparisons in experiments will not be influenced. In general, PDA-LDA outperforms LDA and AS-LDA regardless of how the number of topic components varies.

Application to Document Classification

As the perplexity does not correlate with human judgments, it may be better to test topic models when using its generated topics for other tasks [Chang *et al.*, 2009]. In this work, we apply LDA, AS-LDA, and PDA-LDA to binary document classification task by utilizing topic distributions learned from them as features for supervised classification models. We should emphasize our goal here is to test the quality of learned topics, but not to compare with state-of-the-art document classification methods. On the other hand, features of topic models can be regarded as a complementary for other supervised methods.

Based on the rating scores associated with the three datasets, we divide all the documents of training, validation, and test sets into two classes. The critical value between for all datasets is specified to be 3.5 ([1-5]). The number of documents in each class are shown in Table 6.

Table 6: Statistics of classification datasets.

Data	1st-class	2nd-class
Yelp	57330	124809
AmazonFood	12486	33567
AmazonSport	2372	5610

We compare LDA, AS-LDA, and PDA-LDA in terms of average accuracy, precision, and F1 metrics. Among them, precision and F1 are first computed in each class and then take weighted average over two classes. We adopt standard supervised classifier, i.e., Support Vector Machine (SVM), in the experiments. As the results shown in Table 5, PDA-LDA achieves best performance across all the datasets. Hence the quality of topics generated by PDA-LDA turns out to be better than those of the other models. Besides, although AS-LDA gains lower perplexity than LDA in Yelp dataset, it does not perform better in terms of classification metrics. This reveals obtaining lower perplexity does not ensure better discriminative power of features for classification performance definitely, which also reveals the robustness of PDA-LDA.

Case Study

To give an intuitive interpretation of PDA-LDA, we randomly sample two users and two items to reveal their topic factors in Figure 3, and show their representative topics (topic index corresponds to the largest value for each user and item topic factor) learned from AmazonFood data in Table 7.

We find that the sampled users and items only focus on several topics, for many values less than 0. Thus, they can be characterized by a few critical topics which have clear indications. For example, user-2 likes chips very much.

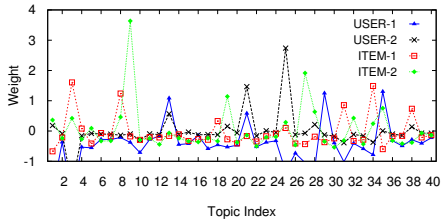


Figure 3: Sampled user and item topic factors.

Table 7: Selected topics from AmazonFood data set.

T3: amazon, price, store, buy, local, order, product, find, bought, shipping
T9: food, cat, cats, chicken, eat, tuna, canned, fish, cans, flavors
T25: chips, flavor, salt, potato, sweet, bag, taste, chip, popchips, good
T27: bars, bar, fat, calories, snack, fiber, grams, protein, healthy, sugar

4 Related Work

Topic models are popularly utilized for modeling text. Different topic models concentrate on different aspects of text. For example, [Blei and McAuliffe, 2007] exploited the labels corresponding to documents to construct their supervised topic model. The authors in [Mei *et al.*, 2008] considered the link relations between documents and assume linked documents have similar topic distributions. Recently, many topic models have been proposed for modeling review text [Titov and McDonald, 2008; Zhao *et al.*, 2010; Sauper *et al.*, 2011; Moghaddam and Ester, 2011; Mukherjee *et al.*, 2014]. Many of them are designed for sentiment analysis, including aspects mining and rating prediction. While our model is not designed specially for review text, but a wider range of text, i.e., user-item connected documents and we concentrate on accounting for the dual role in the topic generation process of the text which are ignored in these models. Collaborative topic model (CTM) [Wang and Blei, 2011] considered the user and item information simultaneously for predicting ratings of users for items. Nevertheless, its topic generation process is the similar to LDA, which does not take the dual role into consideration. Author-topic model [Rosen-Zvi *et al.*, 2004] is somewhat similar with our model as it can obtain the topic

distribution of users. However, the topic generation is only influenced by users and it still cannot model the dual role phenomenon. Many papers deal with dual roles in different research fields, such as the dual role of hashtag in social networks [Yang *et al.*, 2012] and users in question answering [Xu *et al.*, 2012]. However, few previous works emphasize the dual role phenomenon in topic generation process of topic models.

5 Conclusion

In this paper, we propose a new probabilistic topic model called PDA-LDA to account for the dual role phenomenon of user-item connected documents. PDA-LDA models the topic generation process through the joint effect of user, item, and background topic factors on the Dirichlet prior of topic distribution. A Gibbs EM based learning algorithm is derived for the new model to learn optimal topic factors. The experimental results on real data collections have shown that PDA-LDA achieves better held-out perplexity and binary classification accuracy than several other models.

Acknowledgments

This work was supported in part by National Basic Research Program of China (973 Program) under Grant No. 2011CB302206, National Natural Science Foundation of China under Grant No. 61272088, and Tsinghua University Initiative Scientific Research Program.

References

[Blei and McAuliffe, 2007] David M. Blei and Jon D. McAuliffe. Supervised topic models. In *in Proc. of NIPS '07, Vancouver, British Columbia, Canada, December 3-6, 2007*, pages 121–128, 2007.

[Blei *et al.*, 2003a] David M. Blei, Thomas L. Griffiths, Michael I. Jordan, and Joshua B. Tenenbaum. Hierarchical topic models and the nested chinese restaurant process. In *in Proc. of NIPS '03 December 8-13, 2003, Vancouver and Whistler, British Columbia, Canada]*, pages 17–24, 2003.

[Blei *et al.*, 2003b] David M. Blei, Andrew Y. Ng, and Michael I. Jordan. Latent dirichlet allocation. *Journal of Machine Learning Research*, 3:993–1022, 2003.

[Chang *et al.*, 2009] Jonathan Chang, Jordan L. Boyd-Graber, Sean Gerrish, Chong Wang, and David M. Blei. Reading tea leaves: How humans interpret topic models. In *in Proc. of NIPS '09. Proceedings of a meeting held 7-10 December 2009, Vancouver, British Columbia, Canada.*, pages 288–296, 2009.

[Griffiths and Steyvers, 2004] Thomas L. Griffiths and Mark Steyvers. Finding scientific topics. *PNAS*, 101(suppl. 1):5228–5235, 2004.

[Heinrich, 2004] Gregor Heinrich. Parameter estimation for text analysis. Technical report, 2004.

[Hofmann, 1999] Thomas Hofmann. Probabilistic latent semantic indexing. In *in Proc. of SIGIR '99, August 15-19, 1999, Berkeley, CA, USA*, pages 50–57, 1999.

[Liu *et al.*, 2013] Bin Liu, Yanjie Fu, Zijun Yao, and Hui Xiong. Learning geographical preferences for point-of-interest recommendation. in Proc. of KDD '13, pages 1043–1051, New York, NY, USA, 2013. ACM.

[McAuley and Leskovec, 2013] Julian J. McAuley and Jure Leskovec. Hidden factors and hidden topics: understanding rating dimensions with review text. In *in Proc. of RecSys '13, Hong Kong, China, October 12-16, 2013*, pages 165–172, 2013.

[Mei *et al.*, 2008] Qiaozhu Mei, Deng Cai, Duo Zhang, and ChengXiang Zhai. Topic modeling with network regularization. In *in Proc. of WWW '08, Beijing, China, April 21-25, 2008*, pages 101–110, 2008.

[Moghaddam and Ester, 2011] Samaneh Moghaddam and Martin Ester. ILDA: interdependent LDA model for learning latent aspects and their ratings from online product reviews. In *in Proc. of SIGIR '11, Beijing, China, July 25-29, 2011*, pages 665–674, 2011.

[Mukherjee *et al.*, 2014] Subhabrata Mukherjee, Gaurab Basu, and Sachindra Joshi. Joint author sentiment topic model. In *in Proc. of SDM '14, Philadelphia, Pennsylvania, USA, April 24-26, 2014*, pages 370–378, 2014.

[Rosen-Zvi *et al.*, 2004] Michal Rosen-Zvi, Thomas L. Griffiths, Mark Steyvers, and Padhraic Smyth. The author-topic model for authors and documents. In *in Proc. of UAI '04, Banff, Canada, July 7-11, 2004*, pages 487–494, 2004.

[Sauper *et al.*, 2011] Christina Sauper, Aria Haghighi, and Regina Barzilay. Content models with attitude. In *in Proc. of ACL '11, 19-24 June, 2011, Portland, Oregon, USA*, pages 350–358, 2011.

[Titov and McDonald, 2008] Ivan Titov and Ryan T. McDonald. Modeling online reviews with multi-grain topic models. In *in Proc. of WWW '08, Beijing, China, April 21-25, 2008*, pages 111–120, 2008.

[Wallach *et al.*, 2009] Hanna M. Wallach, David M. Mimno, and Andrew McCallum. Rethinking LDA: why priors matter. In *in Proc. of NIPS '09. December 7-10, 2009, Vancouver, British Columbia, Canada.*, pages 1973–1981, 2009.

[Wallach, 2006] Hanna M. Wallach. Topic modeling: Beyond bag-of-words. ICML '06, pages 977–984, New York, NY, USA, 2006. ACM.

[Wang and Blei, 2011] Chong Wang and David M. Blei. Collaborative topic modeling for recommending scientific articles. In *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Diego, CA, USA, August 21-24, 2011*, pages 448–456, 2011.

[Xu *et al.*, 2012] Fei Xu, Zongcheng Ji, and Bin Wang. Dual role model for question recommendation in community question answering. In *in Proc. of SIGIR '12, Portland, OR, USA, August 12-16, 2012*, pages 771–780, 2012.

[Yang *et al.*, 2012] Lei Yang, Tao Sun, Ming Zhang, and Qiaozhu Mei. We know what @you #tag: does the dual role affect hashtag adoption? In *in Proc. of WWW '12, Lyon, France, April 16-20, 2012*, pages 261–270, 2012.

[Yin and Wang, 2014] Jianhua Yin and Jianyong Wang. A dirichlet multinomial mixture model-based approach for short text clustering. In *in Proc. of KDD '14, New York, NY, USA - August 24 - 27, 2014*, pages 233–242, 2014.

[Zhao *et al.*, 2010] Wayne Xin Zhao, Jing Jiang, Hongfei Yan, and Xiaoming Li. Jointly modeling aspects and opinions with a maxent-lda hybrid. In *in Proc. of EMNLP '10, 9-11 October 2010, MIT Stata Center, Massachusetts, USA, A meeting of SIGDAT, a Special Interest Group of the ACL*, pages 56–65, 2010.