Exploiting User Consuming Behavior for Effective Item Tagging

Shen Liu
School of Economics and Management
Tsinghua University
Beijing, China 100084
liush.15@sem.tsinghua.edu.cn

Hongyan Liu*
School of Economics and Management
Tsinghua University
Beijing, China 100084
hyliu@tsinghua.edu.cn

ABSTRACT

Automatic tagging techniques are important for many applications such as searching and recommendation, which has attracted many researchers' attention in recent years. Existing methods mainly rely on users' tagging behavior or items' content information for tagging, yet users' consuming behavior is ignored. In this paper, we propose to leverage such information and introduce a probabilistic model called joint-tagging LDA to improve tagging accuracy. An effective algorithm based on Zero-Order Collapsed Variational Bayes is developed. Experiments conducted on a real dataset demonstrate that joint-tagging LDA outperforms existing competing methods.

CCS CONCEPTS

• **Information systems** → **Collaborative and social computing systems and tools**; **Social tagging systems**;

KEYWORDS

tag recommendation; generative model; user behavior modeling

1 INTRODUCTION

Tagging is a widely adopted approach to organize and use online resources effectively. Various items on the Internet are associated with relevant keywords to help perform navigation or exploration. Such items include user-generated contents like pictures or documents, and products or services in scenarios such as e-commerce.

Considering the tremendous volume and dynamic nature of online resources, manually tagging is insufficient and we resort to automatic tagging for help. Many platforms support social tagging, where every user can annotate item with any keywords. Some research works rely on users' tagging behavior to perform tagging. Such methods suffer from the fact that in many real application scenarios, most user rarely participate in tagging. Some other methods focus on leveraging items' content information for predicting their tags. These content-based methods are mainly used for annotating digital contents like documents and pictures, and their performance are therefore affected by the quality of contents.

Our major improvement over existing works is to use users' consuming behavior for tagging. By consuming behavior, we mean exploratory interactions related to online consuming, such as searching, browsing and purchasing or downloading. Historical data of consuming behavior are quite abundant, and obviously such data imply some relations among items. A key observation is that users tend to browse similar items in a short consecutive time period, which is called a session. Such observation also agrees with some marketing and psychology research works: in a session, consumers may perform intensive exploratory behavior among similar items, so that they can build the knowledge and make better choices. In this paper, those items being visited by the same user in the same session are referred as *neighbors*. Intuitively, items that are frequently observed as neighbors should share some common characteristics, which could be used to perform tagging. In this sense, we can regard user's consuming behavior as implicit collaborative tagging behavior. Yet considering the availability of data, its applicability is wider than explicit collaborative tagging. In many cases, items' content information like short pieces of description or relevant pictures is also easy to collect, we can improve tagging quality by leveraging both sources of information. Therefore in this paper, we propose a probabilistic generative model called Joint-Tagging LDA that can perform tagging with multi-source input.

The main contributions of this paper include:

- We propose an alternative way to perform collaborative tagging in absence of explicit tagging behavior. In this way, automatic tagging can be applied to more scenarios.
- To simultaneously leverage consuming behavior and content information for item annotation, we propose the Joint-Tagging LDA model and an effective inference algorithm to learn the model.
- The performance of Joint-Tagging LDA is evaluated on a real world dataset. The experimental result shows that exploiting consuming behavior for tagging is both feasible and effective.

The rest of the paper is organized as follows: Section 2 presents the proposed Joint-Tagging LDA and its extension. Corresponding inference algorithm is also introduced in this section. In section 3, we evaluate the proposed model and compare it with some competing methods. We review related works in section 4 and conclude our work in section 5.

2 MODEL DESCRIPTION

2.1 from Consuming Behavior to Collaborative Tagging

The core idea of leveraging users' consuming behavior for collaborative tagging is inspired by previous research works. Some

*Corresponding author.

CIKM'17 , November 6–10, 2017, Singapore, Singapore

ACM ISBN 978-1-4503-4918-5/17/11…$15.00
https://doi.org/10.1145/3132847.3133071

Table 1: Symbols and Notations

Symbol	Definition and Description
D	Total number of items
K	Total number of topics
V	Total number of unique terms
T	Total number of unique tags
N_i	Number of terms/neighbors of item i
M_i	Number of tags of item i
w	content terms or neighbors
l	tags
(c),(n)	superscripts denoting terms/neighbors
α, β, γ, η	priors of item topics/terms/neighbors/tags
$\boldsymbol{\theta}$, $\boldsymbol{\phi}$, $\boldsymbol{\psi}$, $\boldsymbol{\varphi}$	distributions of item topics/terms/neighbors/tags

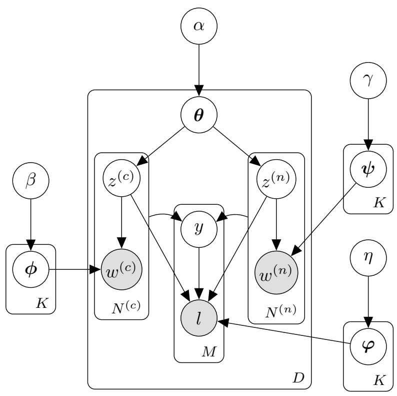


Figure 1: Graphical Model of Joint-Tagging LDA

marketing research works classify users' online shopping strategies into four categories: directed buying, searching/deliberation, knowledge building and hedonic browsing. Every user performs these strategies in each consuming session, and three of them, except the last one, result in low category variety of products in a session[8]. Recent research then showed that users tended to perform intensive exploratory behavior in a session[2]. These facts are also in accordance with our daily experience: during consuming sessions, we usually conduct exploration of similar items, compare them and finally choose one or some of them.

Based on these works and observations, we can infer that consuming behavior in a session implies some topical similarities among items, hence helpful for tagging. Each user's consuming behavior can be modeled as a sequence of sessions, and each session consists of items explored by the user. Then, for every item i, we traverse all users' consuming sequences and find items which co-occur with item i. These items explored in the same session are defined as *neighbors*. Neighbors of item i, together with its content information, can be jointly used to annotate item i. In a sense, we could say all users collaboratively help perform tagging implicitly through their consuming behaviors.

2.2 Joint-Tagging LDA

Joint-tagging LDA is a unified model where neighbors and content information are simultaneously devoted to item annotation. Content information can be regarded as a bag of terms, and some of these terms can indicate tags. Moreover, neighbors of an item, especially frequent ones, are also good tag indicators. We assume that every item can be projected on a simplex in the topic space, where every topic has its own distribution over tags, terms and neighbors. Out of this consideration, Joint-tagging LDA annotates items in the following way: for every item, the algorithm firstly assigns topics to terms/neighbors according to item's topic distribution; then it finds proper tag indicators among terms and neighbors; finally tags are allocated based on topics of these indicators.

Figure 1 shows the graphical model of Joint-Tagging LDA. For every neighbor of item i, its topic $z^{(n)}$ is firstly sampled from the topic distribution θ_i; then a specific neighbor is sampled conditioned on this topic. Content terms are sampled in a similar way. Such sampling process samples $N_i^{(c)}$ times for content term and $N_i^{(n)}$ times for neighbors. After that, a selection variable y is uniformly sampled from $1, \ldots, N_i^{(c)} + N_n^{(n)}$ for M times, where M is the number of tags item i has. Variable y determines which neighbors or terms can be representatives of those tags: for each tag l_m, its topic is the same as the topic of its representative, and then the tag is sampled conditioned on this topic. The generative process of Joint-Tagging LDA can be expressed as follows:

(1) For each topic $k \in 1, \ldots, K$, sample $\phi_k \sim Dir(\beta)$, $\psi_k \sim Dir(\gamma)$ and $\varphi_k \sim Dir(\eta)$
(2) For each item $i \in 1, \ldots, D$:
  (a) Sample its topic distribution $\theta_i \sim Dir(\alpha)$
  (b) For each term position j of item i:
    (i) Sample the term topic $z_n^{(c)} \sim Multi(\theta_i)$
    (ii) Sample a term $w_n^{(c)} \sim Multi(\phi_{z_n^{(c)}})$
  (c) For each neighbor position j of item i:
    (i) Sample the neighbor topic $z_n^{(n)} \sim Multi(\theta_i)$
    (ii) Sample a term $w_n^{(n)} \sim Multi(\psi_{z_n^{(n)}})$
  (d) For each tag position t of item i:
    (i) Sample $y \sim Unif(1, \ldots, N_i^{(c)} + N_i^{(n)})$
    (ii) If $y \le N_d$, sample tag $l_m \sim Multi(\varphi_{z_y^{(c)}})$, else sample tag $l_m \sim Multi(\varphi_{z^{(n)}_{y-N_i^{(c)}}})$

For a certain item, not all of its neighbors can indicate its tags. The same fact also applies to content terms. Through the selection variable y, Joint-Tagging LDA can balance the indicative effect of different resources and select proper representatives automatically.

The inference is intractable for topic model. To deal with it, researchers have developed several algorithms for estimating posteriors. Among these methods, Zero-Order Collapsed Variational Bayes (CVB0) has been justified to have the best performance on both efficiency and stability[1]. So we infer our model following the framework of CVB0. The details are omitted for brevity.

Given a new item, we can infer its topic distribution $\boldsymbol{\theta}$ with fixed term distribution$\boldsymbol{\phi}$ and neighbor distribution $\boldsymbol{\psi}$. The predicted probability and the rank of each tag are then evaluated by calculating $\boldsymbol{\theta\varphi}$, where $\boldsymbol{\varphi}$ is the estimated tag distribution.

3 EXPERIMENTS AND RESULTS

3.1 Experiment Setups

The dataset used for experiments is collected from an Android app store operated by a Chinese Internet Company. Totally 15,068 apps are selected for experiments, and these apps are browsed for at least 100 times by 100,000 users in three months. Tags of these apps are relatively stable as they are annotated manually by employees of the company. Then we filter out tags being assigned to less than

Table 2: Comparison of Different Methods

	Recall@P			NDCG@P		
	P=3	P=5	P=8	P=3	P=5	P=8
L-LDA	0.2010	0.2316	0.2600	0.2558	0.2517	0.2598
LinkLDA	0.4514	0.5589	0.6474	0.5102	0.5394	0.5724
corrLDA	0.4686	0.5727	0.6611	0.5194	0.5437	0.5775
JT-LDA	0.5686	0.6853	0.7742	0.6385	0.6610	0.6927

five of these apps and get 931 unique tags. Averagely 2.37 tags are assigned to each app. Besides description texts of these apps, we also collect 20,000 frequent users' interaction sequences (i.e. browsing and downloading sequences). After that, we extract neighbors of each item with a time gap threshold of 3 minutes. 1,500 apps are randomly selected as test data and the rest are used for training.

We compare the performance of Joint-Tagging LDA with the following tagging methods: Labeled-LDA[10], Link-LDA[5] and Corr-LDA[3]. These methods are based on topic models, and they have been used for annotation in many research works[4, 12, 14]. In all these works, items are annotated based on their content information, so here we use description texts of apps as input for these models. Each of these models will output a ranked list of tags for each app, where the high-ranked tags are more likely to be assigned. Prior parameters of these models as well as Joint-Tagging LDA are set as: $\alpha = 10/K$, and $\beta = \gamma = \eta = 0.001$.

Two evaluation metrics are adopted to compare these models: recall@P and NDCG@P. Reasons for using such metrics are twofolds: firstly, the ground truth, *i.e.* tags manually assigned to apps, can be relatively accurate yet not inclusive, so it is more appropriate to evaluate models based on occurrence times of ground truth tags in their output lists; secondly, the output lists are ranked, thus we should also evaluate the ranking quality. Given an output list of tags $\boldsymbol{r}$ and the ground truth tags $\boldsymbol{y}$, we define $y_{r_l} = 1$ if the l-th tag in $\boldsymbol{r}$ appears in $\boldsymbol{y}$ and $y_{r_l} = 0$ otherwise. The two metrics are calculated as follows:

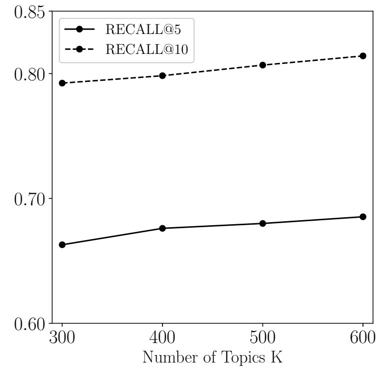
$$Recall@P(\boldsymbol{r}, \boldsymbol{y}) = \frac{\sum_{l=1}^{P} y_{r_l}}{P}$$

$$NDCG@P(\boldsymbol{r}, \boldsymbol{y}) = \frac{\sum_{l=1}^{P} \frac{y_{r_l}}{log(1+l)}}{\sum_{l=1}^{P} \frac{1}{log(1+l)}}$$

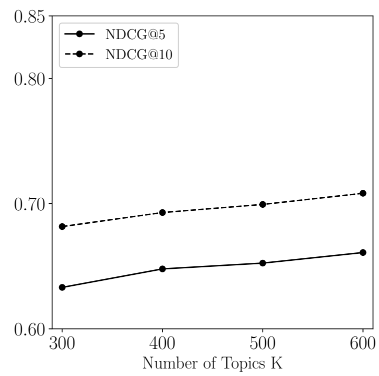
From the formulas we can see, NDCG@P takes the relative ranking of output lists into account while Recall@P does not consider the ranking.

3.2 Experiment Results

Effect of Topic Number K: Figure 2 shows the performance of Joint-Tagging LDA improves as the topic number K increases. By estimating topic-tag distribution $\boldsymbol{\varphi}$, tags are clustered according to their underlying semantics and connections. Yet the distribution of clusters is not balanced, and a larger K can help reveal some small clusters, so that it improves the tagging quality. For example, the tag *breakout bricks* is assigned to several games. In the case $K = 600$, an individual topic associated with this tag appears. Top 5 frequent tags and corresponding probabilities of this topic are: *breakout bricks*:0.2085, *casual*:0.1213, *standalone*:0.0757, *interactive*:0.0738 and *puzzle*:0.0654. However, when $K = 300$, such topic



(a) Recall@P (P=5,10)



(b) NDCG@P (P=5,10)

Figure 2: Joint-Tagging LDA with different topic number K

is not distinguished from other topics that are related to casual games. For other topic models in our experiment, topic number K has similar effect.

Comparison of Different Methods: In the following experiment, we set topic number $K = 600$ for all these models. Table 2 illustrates the superiority of Joint-Tagging LDA (JT-LDA in the table). As we can see, Joint-Tagging LDA always outperforms content-based methods. Out of all the methods being compared, Labeled-LDA gives the worst result. It is assumed that one tag associates with exactly one topic in Labeled-LDA. However, tags in our dataset can be strongly related, and some of them are assigned to relatively few apps. Labeled-LDA can not make stable estimation in this case.

Case Study: In Table 3, we list some apps whose tags largely vary from the tagging results of Joint-Tagging LDA. From the table we can find that some of these apps are not well annotated before. Existing tags can be both insufficient and inconsistent, while our method provides much better tags. On the other hand, for some apps, generated tags are not good. The last line of Table 3 gives such an example. *Guide for Front Line* is a game guidance app. Joint-Tagging LDA assigns some tags related to its content, such as *guns* and *shoot*, but misses an important tag *guidance*. And the tag *game* is related to this app, but not very accurate, which might be attribute to the fact that this app is usually browsed together with some game type apps. We will further study this phenomenon in

Table 3: Case Study: Some Miss-tagging Apps by Joint-Tagging LDA*

App Name	Introduction	Origin Tags	Top 3 Recommended Tags by JT-LDA
Classical Article Today	A reading app that collect articles from social medias	ring-tone	e-book, entertainment, utilities
Pencil Sketch	A pencil sketch tutorial app	tutorial	drawing, sketch, learning
Where Are U	An app aiding in locating the young and elderly	tools	utilities, life, navigation
Minsheng Easyloan	A personal loan app of Minsheng Company	productivity	bank, pay, financing
Shanghai Commuting	A travel app for inquiring traffics in Shanghai	office	trip, life, traffic inquiry
Guide for *Front Line*	A guidance for a first-person shooter game	guidance, auxiliaries	shoot, guns, game

* Original apps are in Chinese. Both tags and app names are translated in English by the author.

our future work. A simple way is to conduct filtering on an item's neighbor based on their category information.

Although we currently only conducted experiments on the app store dataset, Joint-Tagging LDA can be applied to other scenarios to annotate various items such as products, services and digital contents, where we can easily collect data of user consuming behavior.

4 RELATED WORKS

In this section, we briefly review related works from information source's perspective and model's perspective.

From the perspective of information source, existing tagging methods can be roughly classified into two categories, i.e. user-based and content-based methods[13, 14]. User-based methods focus on exploring the historical tagging behaviors of users. Some of these works are analogous to collaborative filtering. For example, Rendle et al. extended matrix factorization to tensor factorization to recommend tags[11]. Because tagging behaviors are relatively rare for most online services, Song et al. argued that these user-based methods could be unstable and inflexible[13]. Content-based methods on the other hand recommend tags for items by their contents[12, 14]. They are believed to be more robust, yet these methods rely on reliable and relevant content information, which limits their applicability. As a rich source of information, user's consuming behavior receives little attention in tagging task and our work fills this gap.

From the perspective of model, our work is most related to topic model and their extensions. Besides models mentioned in experiments, Iwata et al.[6] extended Corr-LDA by considering the relevance between tags and content; Lu et al.[7] thought in social tagging, some tags were subjective and related to users' opinions on items, so they incorporated users' preferences in their model. Many other algorithms like multi-label classification algorithms[9] can also be used for tagging, but they rely heavily on feature engineering.

5 CONCLUSIONS

In this paper, we propose to use user consuming behavior for implicit collaborative tagging. Such method has a wider applicability, as it can be used in scenarios where social tagging behavior is limited. We also propose a novel topic model named Joint-Tagging LDA for automatic tagging with both content information and consuming behavior as input. Experiments conducting on a real dataset shows our model can lead to nearly 10% improvement over competing methods.

ACKNOWLEDGMENTS

This work was supported in part by National Natural Science Foundation of China under grant numbers: 71771131, 71272029 and 71490724, and the Beijing Municipal Natural Science Foundation under grant No. 4152026.

REFERENCES

[1] Arthur Asuncion, Max Welling, Padhraic Smyth, and Yee Whye Teh. 2009. On smoothing and inference for topic models. In *Proceedings of the Twenty-Fifth Conference on Uncertainty in Artificial Intelligence.* AUAI Press, 27–34.

[2] Austin R Benson, Ravi Kumar, and Andrew Tomkins. 2016. Modeling User Consumption Sequences. In *Proceedings of the 25th International Conference on World Wide Web.* International World Wide Web Conferences Steering Committee, 519–529.

[3] David M Blei and Michael I Jordan. 2003. Modeling annotated data. In *Proceedings of the 26th annual international ACM SIGIR conference on Research and development in informaion retrieval.* ACM, 127–134.

[4] Markus Bundschus, Shipeng Yu, Volker Tresp, Achim Rettinger, Mathaeus Dejori, and Hans-Peter Kriegel. 2009. Hierarchical bayesian models for collaborative tagging systems. In *Data Mining, 2009. ICDM'09. Ninth IEEE International Conference on.* IEEE, 728–733.

[5] Elena Erosheva, Stephen Fienberg, and John Lafferty. 2004. Mixed-membership models of scientific publications. *Proceedings of the National Academy of Sciences* 101, suppl 1 (2004), 5220–5227.

[6] Tomoharu Iwata, Takeshi Yamada, and Naonori Ueda. 2009. Modeling social annotation data with content relevance using a topic model. In *Advances in Neural Information Processing Systems.* 835–843.

[7] Caimei Lu, Xiaohua Hu, Xin Chen, Jung-Ran Park, TingTing He, and Zhoujun Li. 2010. The topic-perspective model for social tagging systems. In *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining.* ACM, 683–692.

[8] Wendy W Moe. 2003. Buying, searching, or browsing: Differentiating between online shoppers using in-store navigational clickstream. *Journal of consumer psychology* 13, 1-2 (2003), 29–39.

[9] Yashoteja Prabhu and Manik Varma. 2014. Fastxml: A fast, accurate and stable tree-classifier for extreme multi-label learning. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining.* ACM, 263–272.

[10] Daniel Ramage, David Hall, Ramesh Nallapati, and Christopher D Manning. 2009. Labeled LDA: A supervised topic model for credit attribution in multi-labeled corpora. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 1-Volume 1.* Association for Computational Linguistics, 248–256.

[11] Steffen Rendle and Lars Schmidt-Thieme. 2010. Pairwise interaction tensor factorization for personalized tag recommendation. In *Proceedings of the third ACM international conference on Web search and data mining.* ACM, 81–90.

[12] Xiance Si and Maosong Sun. 2009. Tag-LDA for scalable real-time tag recommendation. *Journal of Computational Information Systems* 6, 1 (2009), 23–31.

[13] Yang Song, Lu Zhang, and C Lee Giles. 2011. Automatic tag recommendation algorithms for social recommender systems. *ACM Transactions on the Web (TWEB)* 5, 1 (2011), 4.

[14] Yong Wu, Yuan Yao, Feng Xu, Hanghang Tong, and Jian Lu. 2016. Tag2Word: Using Tags to Generate Words for Content Based Tag Recommendation. In *Proceedings of the 25th ACM International on Conference on Information and Knowledge Management.* ACM, 2287–2292.