On Modeling Brand Preferences in Item Adoptions

Minh-Duc Luu, Ee-Peng Lim and **Freddy Chong Tat Chua**
{mdluu.2011, eplim, freddy.chua.2009} @smu.edu.sg
Living Analytics Research Centre, School of Information Systems
Singapore Management University, 80 Stamford Road, Singapore 178902

Abstract

In marketing and advertising, developing and managing brands value represent the core activities performed by companies. Successful brands attract buyers and adopters, which in turn increase the companies' value. Given a set of user-item adoption data, can we infer brand effects from users adopting items? To answer this question, we develop the Brand Item Topic Model (BITM) that incorporates users' brand preferences in the process of item adoption by the users. We evaluate our model using synthetic and two real world datasets against baseline models which do not consider brand effects. The results show that BITM can determine users who demonstrate brand preferences and predict item adoptions more accurately.

1 Introduction

1.1 Motivation

Our behaviors of adopting items are complex. They are determined by a number of factors including personal interests, features of items to be adopted, budget constraints and, in most of the cases, by *brands* of the items. In this paper, we focus on item adoptions that can be attributed to user preference of *item topics* and *item brands*. Item topic refers to the latent item group an item can be associated with. Item brand refers to the person, trademark, or business that the item can be identified with so as to make the item distinguishable from other items. For example, the topic of a KFC restaurant is fast food while KFC is a brand. In the case of movie, brand may refer to the director and lead actors in a movie.

In marketing and advertising, developing and managing brand value represent the core activities. A successful brand attracts buyers and adopters, which in turn increases the company's value. An important and relevant question to ask is: *can we model brand effect in user adopting items?* The answer to this question is important for a number of reasons. Firstly, we want to know if a brand is preferred by users who care about brand. If so, it will be easier to sell items under the brand and the marketing strategy for such items may focus more on highlighting the brand rather than the item features.

Secondly, we want to know if a user actually adopts items based on brand, i.e., is *brand conscious*. A user is considered to be brand conscious if a large portion of his/her adoption decisions is based on brands. These brand-conscious users are likely to be more loyal. They are also more willing to pay or advocate more for items of a preferred brand. Once a user is known to be brand conscious, he may be the prime target of marketing and advertising for the reason that he can bring more value to a brand.

1.2 Objective and Contributions

To answer the above important questions, we propose a graphical model called **Brand Item Topic Model** (**BITM**). BITM extends the well known Latent Dirichlet Allocation (LDA) model by modeling brands and decisions of item adoptions, in addition to modeling the latent topics of users and items. The major goal is thus to discover latent variables (topic, brand and decision) from the observed item adoption data.

BITM, to the best of our knowledge, is the first topic model that explicitly examines the brand factor in users' adoption. There are several research challenges in formulating this model. Firstly, we do not assume the price tags of items to be available. We thus cannot determine exclusive brands simply by their high price items compared with the price of similar items under other brands. Not relying on price information however allows us to design a model that can be applied even when price information is not available or when the assumption of exclusive brands having expensive items is not true.

Secondly, it is not trivial to evaluate models that infer brand effects from item adoption data due to a lack of ground truth data. It is possible to solicit user input about their item adoption decisions but this evaluation approach has several drawbacks. It is clearly not scalable. Either users find it intrusive or they may not recall their adoption decisions. We thus have to evaluate BITM using alternative approaches.

This paper addresses the above challenges by offering the following contributions.

- We propose a novel topic model BITM for inferring brand and topic latent variables that generate a set of observed item adoptions without price information. The model introduces a richer structure of adoption process.

- To evaluate BITM, we generate a synthetic dataset where *exclusive* brands and *brand conscious* users are injected and controlled by a set of parameters. We show that BITM outperforms baselines in learning the ground truth variables and it also achieves reasonable accuracy in recovering brand conscious users and exclusive brands.
- We also evaluate BITM using two real datasets from FourSquare and ACM Digital Library. The exclusive food outlet brands learnt from the Foursquare data by BITM are shown to be more pricey than the non-exclusive ones. We also show that the exclusive authors learnt from the latter dataset have higher h-index than those non-exclusive ones.

1.3 Paper Outline

The rest of this paper is organized as follows. We cover the related works in Section 2. Our proposed BITM model and its learning formulae are given in Section 3. The evaluations of BITM on synthetic data and real data are covered in Sections 4 and 5 respectively. We also evaluate the ability of BITM predicting item adoptions in Section 6. Finally, we conclude in Section 7.

2 Related Works

2.1 Item Adoption using Latent Factors

When modeling users adopting items, most of the previous research focused on modeling the user and item latent factors corresponding to item topics, item features and user preferences that generate the adoptions (Hofmann 2001), (Hofmann 2003), (Hofmann 2004), (Liu, Zhao, and Yang 2009). Jin et al. performed an empirical study and showed that the rating patterns of users could be separated from the users preferences (Jin, Si, and Zhai 2002). For example, among users with similar preferences, some users tend to give higher ratings than others. This suggests that there is an alternative process in how users like or adopt items. In our work, we study such an alternative process by assuming that users tend to adopt based on the brand factor.

In our work, we also represent user's topic preferences by distributions i.e. different topics occupy a user's interest with different proportions. However, we differentiate our BITM model from these prior works by introducing a different process in how users adopt items. This process assumes that users adopt items not only because of topic preferences but also their preference for certain brands. Usually, for an item co-created by multiple brands, only a small subset of the brands are considered in a user's adoption decision. In this paper, we assume that he will use *only one* brand for the decision. Generally, for an item associated with multiple brands, it is interesting but challenging to discover the actual *latent brand* from which a user chooses to adopt the item. We design BITM model such that it will recover the actual latent brands in brand-based adoption decisions.

2.2 Role of brands in adoption process

In traditional offline markets, the importance of brands, more precisely brand loyalty, has been recognized in marketing literature in numerous studies such as (Jacoby and Chestnut 1978), (Dick and Basu 1994), (Chaudhuri and Holbrook 2001), (Aaker 2012). Brands play a major role in adoption process since they help to shape consumer perceptions and tastes, thus inspire adoption behavior. In this way, strong brands not only create demand but also continuity of demand into the future by leveraging consumer's loyalty to the brands. Moreover, (Dick and Basu 1994), (Aaker 2012) noted that brand loyalty leads to other marketing advantages such as favorable word of mouth, reduced marketing cost. These studies confirm the importance of determining strong brands and users who are loyal to those brands; which justifies contribution of our BITM model.

In online markets, brands continue to prove their importance (Erdem and Keane 1996), (Erdem et al. 1999), (Smith and Brynjolfsson 2001). By comparing the effects of pricing in online shopbots (i.e. internet services that compare prices of similar consumer goods sold on different online websites), Smith and Brynjolfsson conclude that higher prices on well-known websites do not affect the sales of products because of the brand effect carried by the well-known websites (Smith and Brynjolfsson 2001). Erdem and Keane performed a temporal analysis of branding effects and found that advertising intensity has only weak short run effects on brand adoption, but has a strong cumulative effect in the long run (Erdem and Keane 1996).

These studies suggest that brand does play an important role in user item adoptions. Our work here combines the body of work on using topic distributions with additional brand latent variables for modeling user adoptions. Such kind of models have not been studied before in the marketing science and recommendation literature.

3 Brand Item Topic Model

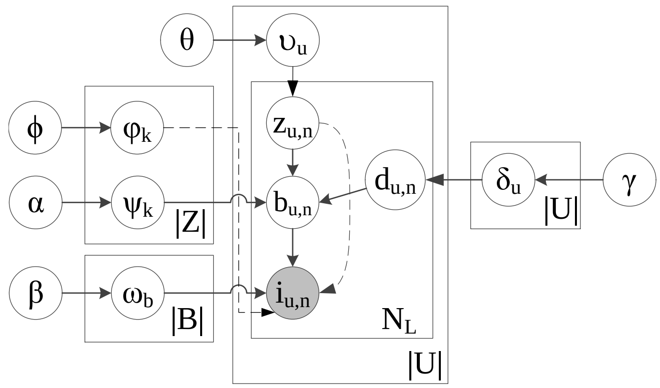
3.1 Topic Model and BITM

Before elaborating our Brand-Item Topic Model, we would like to briefly present a well-known topic model, Latent Dirichlet Allocation (LDA). LDA is a generative Bayesian model in which each document of a collection is modelled as a finite mixture over an underlying set of topics (Blei, Ng, and Jordan 2003). Each document is associated with a *topic distribution*. Each topic is in turn modeled as a *distribution over terms*. In the context of item adoption, LDA can be adapted as follows.

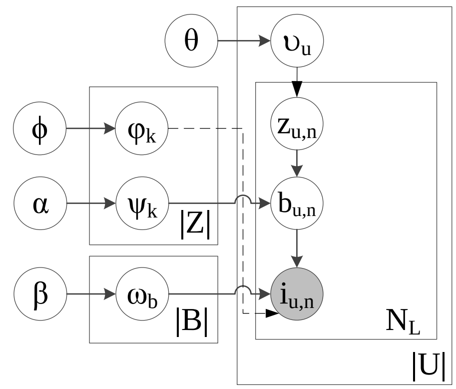
- Each topic can be considered as a *category* of items. Thus, a topic is characterized by a distribution over *items*.
- Each user is modelled as a "document" e.g. we can model his/her preference as a topic distribution. More precisely, the adoption history of a user can be considered as a "document of adopted items" where each item is generated under some favorite topic of the user.

Our proposed model, Brand-Item Topic Model (BITM), incorporates brand preference into the topic model. To choose an item for adoption, a user first chooses a topic z that she is interested in. Then the user chooses the item to adopt based on either one of the following methods:

1. *Topic-based Adoption*: With the chosen topic z, the user adopts one of the many items under z. For example, when



(a) Brand-Item Topic Model (BITM)



(b) Simplified Brand-Item Topic Model (sBITM)

Figure 1: Bayesian network for BITM and sBITM

the user buys a novel, a topic-based adoption will have the user first selects a genre (e.g. sci-fi) among his topic preferences and selects a novel based on its popularity under the genre.

2. *Brand-based Adoption*: The user chooses a brand b from topic z, then selects an item from the item distribution of brand b. For example, when user buys a novel, the user first selects a genre, then a preferred author under the selected genre, follow by selecting a novel written by the preferred author.

It can be seen that for a given user, the choice of method for adoption reveals the importance of brand to her. If she usually prefers to adopt based on popular brands, we may say that she has a *brand-preference*. Understanding users' brand preferences allows us to utilize more information for making better recommendations to them.

3.2 Generative Process

We illustrate the generative process of our Brand-Item Topic Model (BITM) using the graphical model in Figure 1a and its corresponding notations in Table 1a. From Figure 1a the generative process is as follows:

1. (Generating distributions from priors):
   (a) For each user u, we sample a multinomial topic distribution $\boldsymbol{\vartheta}_u$ and a decision distribution $\boldsymbol{\delta}_u$.
   $$\boldsymbol{\vartheta}_u \sim Dir(.|\theta), \quad \boldsymbol{\delta}_u \sim Dir(.|\gamma)$$
   (b) For each topic k, we sample an item distribution $\boldsymbol{\varphi}_k$ and a brand distribution $\boldsymbol{\psi}_k$. For each topic, its brand distribution represents how exclusive each brand is. More exclusive brands should have larger values.
   $$\boldsymbol{\varphi}_k \sim Dir(.|\phi), \quad \boldsymbol{\psi}_k \sim Dir(.|\alpha)$$
   (c) For each brand b, we sample an item distribution $\boldsymbol{\omega}_b$ from Dirichlet distribution with prior parameter β.
   $$\boldsymbol{\omega}_b \sim Dir(.|\beta)$$
2. (Generating adoptions):
   (a) User u makes n-th item adoption by first sampling a topic $z_{u,n} = k$ from his topic distribution $\boldsymbol{\vartheta}_u$,
   $$z_{u,n} = k \sim Multi(\boldsymbol{\vartheta}_u)$$
   (b) User u then decides whether this adoption should be based on topic or brand using a decision variable $d_{u,n}$, sampled from her decision distribution $\boldsymbol{\delta}_u$.
   $$d_{u,n} \sim Bernoulli(\boldsymbol{\delta}_u)$$
   (c) User u then either adopts item v from the topic-item distribution $\boldsymbol{\varphi}_k$ if $d_{u,n} = \mathtt{T}$ or chooses a brand b from the topic-brand distribution $\boldsymbol{\psi}_k$ if $d_{u,n} = \mathtt{B}$. If a brand is chosen, item v is sampled from the brand-item distribution $\boldsymbol{\omega}_b$.
   $$i_{u,n} = v \sim \begin{cases} \boldsymbol{\varphi}_k & \text{when } d_{u,n} = \mathtt{T} \\ \boldsymbol{\omega}_b & \text{when } d_{u,n} = \mathtt{B} \end{cases}$$
   $$b \sim Multi(\boldsymbol{\psi}_k) \quad \text{if } d_{u,n} = \mathtt{T}$$

We would like to note some important features of our model.

- All Dirichlet priors are symmetric.
- The brand-item distribution w_b is not observable from data since the brand b is *latent*.

It can also be seen that when all adoption decisions are topic-based, the BITM generative process degenerates to that of LDA. Hence, LDA can be considered as a special case of BITM. On the other hand, if all decisions are brand-based, we have a simplified version of BITM, denoted as sBITM (see Figure 1b), which can be used as another baseline in evaluating BITM.

Training BITM requires extensive computation cost as we have five distributions to be learned. Thus, instead of using variational methods (Blei, Ng, and Jordan 2003), (Hoffman, Bach, and Blei 2010), and the Expectation-Maximization algorithm (Dempster, Laird, and Rubin 1977), (Redner and Walker 1984) we adopt Gibbs sampling (Gelfand and Smith 1990), (Casella and George 1992).

Notation	Description
$i_{u,n}$	Item at n-th adoption of user u
$z_{u,n}$	Latent topic of $i_{u,n}$
$b_{u,n}$	Latent brand of $i_{u,n}$
$d_{u,n}$	Latent decision variable of this adoption
$\boldsymbol{\omega}_b$	Parameters for the item distribution of brand b
β	Hyperparameter for Dirichlet prior of $\boldsymbol{\omega}_b$
$\boldsymbol{\psi}_k$	Parameters for the brand distribution of topic k
α	Hyperparameters for Dirichlet prior of $\boldsymbol{\psi}_k$
$\boldsymbol{\vartheta}_u$	Parameters for the topic distribution of user u
θ	Hyperparameters for Dirichlet prior of $\boldsymbol{\vartheta}_u$
$\boldsymbol{\varphi}_z$	Parameters for the title distribution of topic z
ϕ	Hyper parameters for Dirichlet prior of $\boldsymbol{\varphi}_z$'s
$\boldsymbol{\delta}_u$	Parameters for binomial distribution of $d_{u,n}$
γ	Hyper parameter for sampling Λ
U, B, Z	Set of users, set of brands and set of topics resp.

(a) For generative process of BITM

Notation	Description
$\boldsymbol{i}_u = \{i_{u,1}, \ldots, i_{u,n}\}$	Vector of adoptions by user u.
$\boldsymbol{z}_u = \{z_{u,1}, \ldots, z_{u,n}\}$	Corresponding latent topics.
$\boldsymbol{b}_u = \{b_{u,1}, \ldots, b_{u,n}\}$	Corresponding latent brands.
$\boldsymbol{d}_u = \{d_{u,1}, \ldots, d_{u,n}\}$	Corresponding latent decisions.
$\mathcal{I} = \{\boldsymbol{i}_1, \ldots, \boldsymbol{i}_N\}$	Vector of item adoptions by all users.
$\mathcal{Z} = \{\boldsymbol{z}_1, \ldots, \boldsymbol{z}_N\}$	Vector of latent topics of adoptions.
$\mathcal{B} = \{\boldsymbol{b}_1, \ldots, \boldsymbol{b}_N\}$	Vector of latent brands of adoptions.
$\mathcal{D} = \{\boldsymbol{d}_1, \ldots, \boldsymbol{d}_N\}$	Vector of latent decisions of adoptions.
$\mathbf{nt}_u = (nt_{u,1}, \ldots, nt_{u,K})$	Topic counts for a user u.
$\mathbf{ni}_k^{tp} = \left(ni_{k,1}^{tp}, \ldots, ni_{k,\lvert I\rvert}^{tp}\right)$	Item counts for adoptions by topic k.
$\mathbf{ni}_b^{br} = \left(ni_{b,1}^{br}, \ldots, ni_{b,\lvert I\rvert}^{br}\right)$	Item counts for adoptions by brand b.
$\mathbf{nb}_k^{br} = \left(nb_{k,1}^{br}, \ldots, nb_{k,\lvert B\rvert}^{br}\right)$	Brand counts for *brand-based* adoptions.
$\mathbf{nd}_u = \left(nd_u^{tp}, nd_u^{br}\right)$	Decision counts for user u.

(b) For learning BITM

Table 1: All notations related to BITM

3.3 Learning latent variables

Using the approach of Gibbs sampling, we can learn latent topics, brands and decisions by an *alternating* update process. We start from an initial guess for the variables and repeat the following update process until convergence.

- Using *current* values of latent decisions $\mathcal{D}_c$, latent topics $\mathcal{Z}_c$ and latent brands $\mathcal{B}_c$, we sample *new* latent decisions $\mathcal{D}_n$.
- Using $\mathcal{D}_n, \mathcal{Z}_c, \mathcal{B}_c$, we sample *new* latent topics $\mathcal{Z}_n$.
- Using $\mathcal{D}_n, \mathcal{Z}_n, \mathcal{B}_c$, we sample *new* latent brands $\mathcal{B}_n$

Due to space constraints, detailed propositions for the update process together with their proofs are not included here. Interested readers can find it in the technical note (Luu 2013a).

3.4 Learning distributions

Once we have learned all latent variables, they can be used to estimate five distributions $\boldsymbol{\vartheta}_u, \boldsymbol{\delta}_u, \boldsymbol{\psi}_k, \boldsymbol{\varphi}_k, \boldsymbol{\omega}_b$ which we are interested in. Similar to LDA, the conjugacy of Dirichlet and Multinomial distributions can be used to show that all the parameters $\boldsymbol{\vartheta}_u, \boldsymbol{\delta}_u, \boldsymbol{\psi}_k, \boldsymbol{\varphi}_k, \boldsymbol{\omega}_b$ follow Dirichlet posteriors. All notations in the proposition follow Table 1b.

Proposition 1 (Learning distributions) *The five interested distributions can be learned as follows.*

1. *Given a user u, his/her topic distribution $\boldsymbol{\vartheta}_u$ and decision distribution $\boldsymbol{\delta}_u$ follow Dirichlet posteriors parameterized by $\theta\mathbf{1} + \mathbf{nt}_u$ and $\gamma\mathbf{1} + \mathbf{nd}_u$ respectively. Thus, we have:*
$$P(\boldsymbol{\vartheta}_u|\mathcal{Z}, \theta) = Dir(\boldsymbol{\vartheta}_u|\theta\mathbf{1} + \mathbf{nt}_u) \quad (1)$$
$$P(\boldsymbol{\delta}_u|\mathcal{D}, \gamma) = Dir(\boldsymbol{\delta}_u|\gamma\mathbf{1} + \mathbf{nd}_u) \quad (2)$$
2. *Given a topic k, its item distribution $\boldsymbol{\varphi}_k$ and brand distribution $\boldsymbol{\psi}_k$ follow Dirichlet posteriors parameterized by $\phi\mathbf{1} + \mathbf{ni}_k^{tp}$ and $\alpha\mathbf{1} + \mathbf{nb}_k^{br}$ respectively. Thus, we have:*
$$P(\boldsymbol{\varphi}_k|\mathcal{I},\ \mathcal{Z}, \mathcal{D},\ \phi) = Dir(\boldsymbol{\varphi}_k|\phi\mathbf{1} + \mathbf{ni}_k^{tp}) \quad (3)$$
$$P(\boldsymbol{\psi}_k|\ \mathcal{I},\ \mathcal{Z}, \mathcal{D}, \alpha) = Dir(\boldsymbol{\psi}_k|\alpha\mathbf{1} + \mathbf{nb}_k^{br}) \quad (4)$$
3. *Given a brand b, its item distribution follows Dirichlet posterior parameterized by $\beta\mathbf{1} + \mathbf{ni}_b^{br}$. Thus, we have*
$$P(\boldsymbol{\omega}_b|\mathcal{I}, \mathcal{B},\ \mathcal{D}, \beta) = Dir(\boldsymbol{\omega}_b|\beta\mathbf{1} + \mathbf{ni}_b^{br}) \quad (5)$$

Since all the parameters follow Dirichlet posteriors, the expectation of Dirichlet posteriors can be used to estimate them. We skip the details here but interested readers can easily find them in any standard reference on Dirichlet posterior e.g. (Heinrich 2005).

4 Experiments on Synthetic Data

To evaluate BITM against two baselines LDA and sBITM, we first conduct experiments using synthetic adoption data that contains ground truth labels, i.e., item's topic label, adopter's brand-consciousness, brand's topic label, and brand's exclusiveness. We also vary the dataset parameters to study how BITM performs under different data settings.

4.1 Synthetic data generation

Table 2: Parameters for Synthetic Data Generation

Symbols	Description	Value Range (Default Value)
N_{user}	# users	10K
N_{brand}	# brands	100
N_{item}	# items	1K
N_{topic}	# topics	$\{5, 10, 15\}$ (10)
R_{adopt}	# adoptions/user	[50,200] (100)
P	% adoptions in favorite topics	90
Q	% brand lovers	[0, 100] (20)
X	% exclusive brands/topic	10

The set of parameters used in synthetic data generation is given in Table 2. For simplicity, every brand is assigned to only one topic. Each item is associated with R_{brand} brands, and thus R_{brand} topics. Every user is assigned K favorite topics and $N_{topic} - K$ non-favorite topics where N_{topic} is the total number of topics. $P\%$ ($P > 50$) of adoptions are reserved for items in the user's favorite topics leaving the remaining $100 - P\%$ to those in non-favorite topics. $Q\%$ of users are brand-conscious and they adopt items based on exclusive brands. $X\%$ of brands for each topic are designated as exclusive brands. We also impose the constraint that each brand has at least 10 items. This ensures enough brand-based adoption data for each brand.

Using the parameters as listed in Table 2, we generate the synthetic data as follows: 1) For each brand, randomly assign a topic label while ensuring that every topic has similar number of brands. 2) For each topic, randomly designate $X\%$ of these brands to be exclusive. 3) For each brand, randomly assign 10 items to ensure that each brand later will have at least 10 items. 4) For each item, randomly assign two brands. 4) For each user, randomly assign two topics as his favorites. 5) Randomly assign $Q\%$ of users to be brand conscious and they will always adopt items of exclusive brands. 6) Every user u is assigned the same number of adoptions R_{adopt} and to generate each adoption of u, first select one of two favorite topics of the user. If he is brand conscious, randomly select an exclusive brand under the topic followed by randomly selecting an item under the exclusive brand. Otherwise (i.e., the user is non brand conscious), randomly select an item under the favorite topic.

4.2 Results

Topic-item distribution error: All three models BITM, sBITM and LDA learn the topic assignments of item adoptions. We aim to evaluate the accuracy of models' topic assignments with respect to the ground truth topic assignments. Since all models are instances of unsupervised learning, we will not be able to exactly recover the ground truth topics after learning. We first have to match the learned topics with the ground truth topics and examine how accurate the matching is. For each model, the matching procedure is described below.

- Given a topic z (either ground-truth or learned), denote its item distribution as $I(z)$. For each ground-truth topic k_t, we first determine k_l, its *best matched* learned topic, to be the one whose item distribution is *closest* to that of the true topic k_t. The closeness is measured by Jensen-Shannon (JS) distance. Thus k_l and the corresponding error in recovering ground-truth topic k_t can be defined as following.

$$k_l := \arg\min_{z_l} JS\left[I(z_l), I(k_t)\right]$$
$$Err(k_t) := JS\left[I(k_l), I(k_t)\right]$$

- By taking average of all $Err(k_t)$, we can define error $TopicErr$ in learning topics for each model as

$$TopicErr := avg_{k_t} Err(k_t) \qquad (6)$$

Figure 2a shows the errors obtained by the three models on learning topic-item distribution where we fix the number of topics as 10 while varying % of brand conscious users Q from 0 to 100. As expected, BITM and LDA produce the same error values when $Q = 0$ whereas BITM and sBITM performs the same when $Q = 100$. As Q increases, both BITM and sBITM show that they can learn the topic labels more accurately when there are more brand conscious users. LDA, on the other hand, generates larger error when Q increases. Finally, the performance of sBITM is worse than LDA if less than 50% of users are brand conscious; which is reasonable since more than 50% of adoption decisions are now topic-based. We also performed paired t-tests (see Table 3) to check if these results are statistically significant. It can be seen that BITM improves significantly over sBITM (LDA) when $Q < 100$ ($Q > 0$) respectively.

Q	t-test	$sBITM$	LDA
0%	$BITM$	2.8E-14*	1.5E-01
25%	$BITM$	6.3E-10*	1.7E-04*
50%	$BITM$	2.0E-09*	1.2E-06*
75%	$BITM$	1.7E-08*	9.2E-08*
100%	$BITM$	6.6E-01	5.7E-12*

Table 3: p values from paired t-tests (2-tail) on errors in learning topics. **Note**: $^*p < 0.01$.

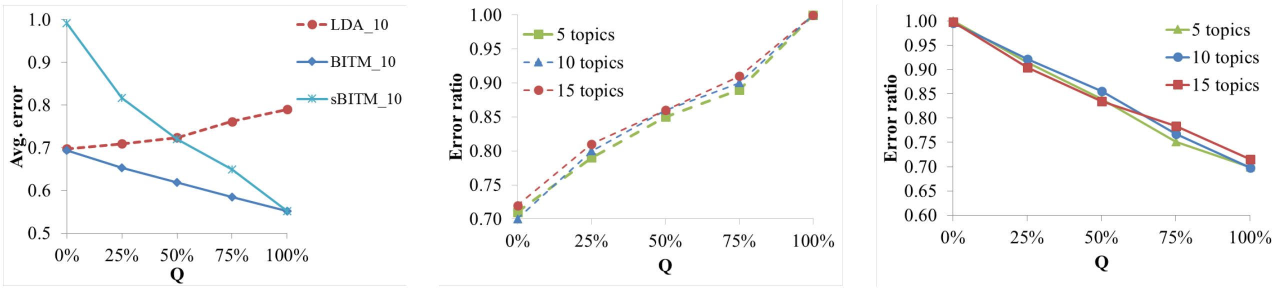
To verify if the improvement by BITM is consistent we vary the number of topics as 5, 10 and 15 respectively. For all settings, we examine topic-item distribution error ratios between BITM and sBITM (Figure 2b); between BITM and LDA (Figure 2c). Figure 2b shows that BITM outperforms sBITM in learning topic-item distribution when $Q < 100\%$ whereas Figure 2c shows that BITM outperforms LDA when $Q > 0$. Moreover, given the same Q, the error ratios for three settings of number of topics are also similar. These results again confirm that BITM is the best among the three models.

Accuracy of brand conscious user prediction: Every user is assigned to be either brand conscious or not brand conscious. We use U_q to denote the set of brand conscious users in the ground truth data, and U'_q to denote the set of brand conscious users learned (or predicted) using BITM. Ideally, we want $U_q = U'_q$. To measure how accurate are the brand conscious users predicted by BITM, we utilize the $Accuracy$ measure as defined:

$$Acc_q(Q) = \frac{|U_q \cap U'_q| + |(U - U_q) \cap (U - U'_q)|}{|U|}$$

Figure 3 shows that the accuracy of predicted brand conscious users improves with increasing $Q\%$. Compared with a random 50-50 guess which has a 0.5 accuracy, BITM can predict brand conscious users quite well with mostly 0.8 accuracy when $Q\%$ is larger than 20%.

Topic-brand distribution error: In the synthetic data, each brand is assigned a ground truth topic. Using the topic-item distributions, we determine the best matched ground truth topic k'_l for each learned topic k_l. Let the topic-brand distribution of k_l and k'_l among item adoptions by brand conscious users be denoted by $A(k_l)$ and $A(k'_l)$ respectively. We define the *topic-brand distribution error* (denoted by $BrandErr$) between learned and ground truth topic-brand distributions using Jensen-Shannon divergence



(a) $TopicErr$'s obtained by three models (# topics = 10)

(b) $\frac{TopicErr_{BITM}}{TopicErr_{sBITM}}$

(c) $\frac{TopicErr_{BITM}}{TopicErr_{LDA}}$

Figure 2: Topic-Item Distribution Errors by various % of Brand Lovers Q

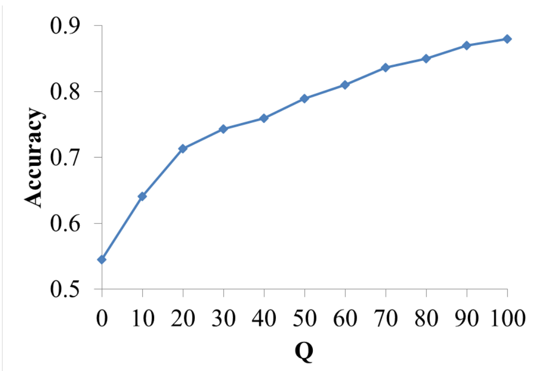


Figure 3: Accuracy of BITM in predicting brand conscious users

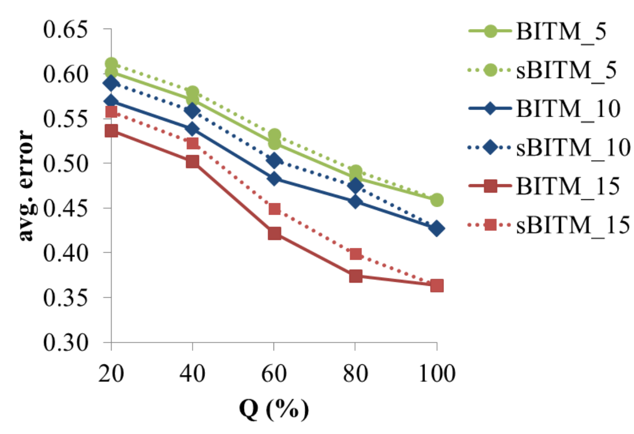


Figure 4: Topic-Brand Distribution Error

measure similar to that for $TopicErr$. As shown in Figure 4, BITM's topic-brand distribution error improves with larger $Q\%$ of brand conscious users and BITM outperforms sBITM when $Q\%$ is less than 100%. LDA is not involved in this evaluation as it does not learn the exclusive brands for each topic.

5 Experiments on Real Data

We conducted a series of experiments on the BITM model using two real world datasets derived from Foursquare and ACM Digital Library (ACMDL) (acm 2011). We first derive subsets of the datasets using a sampling strategy that trims away users with very few adoptions. The experiments then seek to uncover the hidden topics and brand preferences in item adoptions using BITM. We also compare the topics derived from BITM with those from LDA.

5.1 Datasets

Our Foursquare dataset consists of check-in data generated by Singapore users from October 2012 to April 2013. Each food outlet is an item, each food outlet chain is a brand and each check-in is an item adoption by a user. From the raw dataset, we selected a subset of the data based using top $k = 100$ brands and denote the selected data as **4SQDB**. The selection steps will be elaborated shortly.

For ACMDL, each citing author is a user, and each publication is an item. Every publication belongs to one or more authors who are also treated as brands in our experiments. Each citation of some publication by some (citing) author is an item adoption. We used publications in ACMDL from 1998 to 2005 to select a subset using top $k = 10$ authors. The selected data is denoted as **ACMDB**.

We explain the steps of choosing for both datasets 4SQDB and ACMDB with the aid of Figure 5. First, we selected top k brands based on the number of item adoptions of the brands. We denote the set of brands as B_1. Then, we selected all items that belong to B_1 and denote this set of items as I_1. Next, we extracted users that have adopted at least one item in I_1. This set of users is denoted by U_0. Then we extracted all other items adopted by U_0. We denote the new set of items as I_2. We extracted all brands of items in $I_0 = I_1 \cup I_2$ and denote this set as B_0. Finally, we filter away users in U_0 with less than two item adoptions, items in I_0 with less than two adoptions from users in U_0, and brands in B_0 that have no items. We repeat this filtering step until all the remaining users, brands and items satisfy the minimum thresholds. We denote the final sets of users, brands and items as U, B, and I. The statistics of two obtained datasets 4SQDB and ACMDB are shown in Table 4.

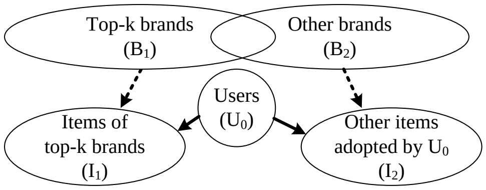


Figure 5: Dataset construction (solid line = adoption, dash line = brand relationship)

Table 4: Data Statistics

Dataset	# users	# brands	# items	# adoptions
4SQDB	5406	622	2444	64,622
ACMDB	356	7520	3790	16,308

5.2 Prior parameters

To determine appropriate prior parameters, we performed grid search and chose optimal parameters which maximizes log likelihood function. After grid search, we got the following values for priors:

$$\alpha = \beta = 0.1; \quad \gamma = 1; \quad \phi = 0.2; \quad \theta = 50/K;$$

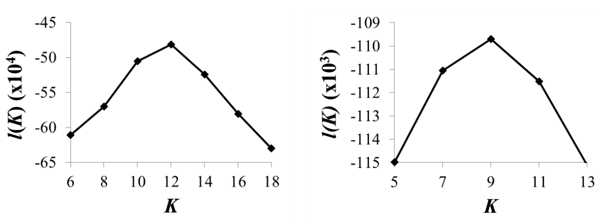
where K is the number of topics.

5.3 Topic Analysis

We first determine the appropriate number of topics for analysing each dataset by running LDA on them. The results in Figure 6 show that the log likelihoods reach maximum at 12 and 9 topics for **4SQDB** and **ACMDB** respectively. Thus, in training models, we empirically used 12 and 9 topics for **4SQDB** and **ACMDB** respectively.

4SQDB: We compare the item distributions of topics discovered by BITM and LDA as shown in Table 5. ALn (ABn) represents the n^{th} topic learnt by LDA (BITM). The similarity between item distributions of two topics is given by the Jensen-Shannon (JS) divergence where smaller JS divergence values indicate higher similarity. From Table 5, we observed that the two models learned quite similar topics as most values in the diagonals of Table 5 are relatively smaller compared to the non-diagonals.

Contrary to our intuition, the learned topics are not about cuisine types (e.g., Chinese food, Indian food) but are clusters of food outlets in 12 different location areas of Singapore as shown in Table 7.



(a) 4SQDB (b) ACMDB

Figure 6: Log likelihood upon training LDA on two datasets. Here $l(K)$ is the log likelihood w.r.t. the number of topics K.

Table 7: Matching learnt topics

Topics (LDA)	Topics (BITM)	Topic Label
4SQDB		
AL1	AB1	Tampines
AL2	AB2	Chua Chu Kang
AL3	AB3	Ang Mo Kio
AL4	AB4	Orchard
AL5	AB5	Pasir Ris
AL6	AB6	Punggol
AL7	AB7	Toa Payoh
AL8	AB8	Hougang
AL9	AB9	Sembawang
AL10	AB10	Jurong
AL11	AB11	Compass Point
AL12	AB12	Bukit Panjang
ACMDB		
TL1	TB1	DB+DM
TL2	TB2	PO
TL3	TB3	SE
TL4	TB4	WWW
TL5	TB5	Systems
TL6	TB6	Security
TL7	TB7	WN
TL8	TB8	CA

ACMDB: We manually determined each topic based on keywords in top-20 titles of that topic. Due to space constraint, the topics discovered and their top-20 paper titles are not provided here but interested readers can find them at extended result (Luu 2013b). We then compare the topics found by BITM and LDA using JS divergence (Table 6). TLn (TBn) represents the n^{th} topic learnt by BITM (LDA) for ACMDB. In Table 6, the columns (rows) show topics learned by LDA (BITM) respectively. Given that smaller JS divergence implies higher similarity, we found that among 9 topics, BITM and LDA agree on 8 topics shown by the bolded diagonal entries of Table 6. These topics are Databases and Data Mining (DB+DM), Power Optimization (PO), Software Engineering (SE), World Wide Web (WWW), System, Security, Wireless Network (WN), Computer Architecture (CA). But LDA discovered the topic information retrieval (IR) which BITM did not. Instead, BITM discovered two sub-topics of software engineering: SE_1 (Algorithms and Programming) and SE_2 (Fault Localization).

5.4 Brand Preference Analysis

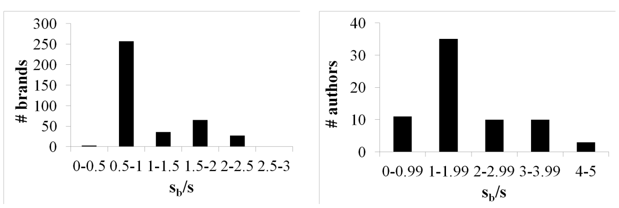
As demonstrated in synthetic experiment, BITM can learn brand preferences of users. It can help to determine whether a user is brand conscious or not. In our experiments, a user is considered as *brand conscious* if at least 80% of his/her adoptions are brand-based. Once the set of brand conscious users is determined, it can be used to identify *exclusive* brands for each topic. This is a major gain provided by BITM as previous models do not help to identify exclusive brands. More specifically, we can identify exclusive brands based on the quantities defined below.

Table 5: **4SQDB**: JS divergence between topic-item distributions learned by BITM and LDA

	AL1	AL2	AL3	AL4	AL5	AL6	AL7	AL8	AL9	AL10	AL11	AL12
AB1	**0.701**	0.915	0.927	0.852	0.940	0.838	0.859	0.756	0.805	0.778	0.728	0.843
AB2	0.967	**0.530**	0.976	0.971	0.671	0.835	0.689	0.974	0.855	0.971	0.947	0.870
AB3	0.934	0.916	**0.675**	0.891	0.870	0.842	0.770	0.955	0.702	0.902	0.905	0.965
AB4	0.933	0.954	0.855	**0.717**	0.941	0.794	0.806	0.905	0.838	0.733	0.924	0.932
AB5	0.883	0.889	0.931	0.910	**0.580**	0.962	0.858	0.937	0.770	0.893	0.881	0.931
AB6	0.877	0.905	0.910	0.903	0.871	**0.516**	0.906	0.893	0.802	0.820	0.816	0.879
AB7	0.889	0.906	0.870	0.877	0.903	0.816	**0.552**	0.878	0.810	0.853	0.874	0.873
AB8	0.935	0.936	0.917	0.950	0.944	0.831	0.794	**0.546**	0.870	0.890	0.944	0.812
AB9	0.885	0.908	0.909	0.924	0.858	0.937	0.829	0.879	**0.611**	0.871	0.854	0.913
AB10	0.958	0.944	0.968	0.917	0.871	0.924	0.819	0.829	0.908	**0.717**	0.945	0.869
AB11	0.894	0.942	0.705	0.949	0.958	0.867	0.807	0.913	0.819	0.964	**0.675**	0.862
AB12	0.898	0.898	0.890	0.914	0.921	0.838	0.841	0.920	0.829	0.928	0.919	**0.697**

Table 6: **ACMDB**: JS divergence between topic-item distributions learned by BITM and LDA

	TL1(DB + DM)	TL2(PO)	TL3(SE)	TL4(WWW)	TL5(System)	TL6(Security)	TL7(WN)	TL8(CA)	TL9(IR)
TB1(DB + DM)	**0.545**	0.937	0.914	0.835	0.914	0.897	0.637	0.899	0.749
TB2(PO)	0.918	**0.354**	0.845	0.867	0.832	0.853	0.902	0.871	0.918
TB3(SE_1)	0.820	0.783	**0.673**	0.795	0.829	0.827	0.844	0.831	0.822
TB4(WWW)	0.862	0.810	0.821	**0.689**	0.804	0.763	0.829	0.810	0.837
TB5(System)	0.852	0.806	0.770	0.843	**0.685**	0.818	0.839	0.829	0.823
TB6(Security)	0.828	0.791	0.815	0.796	0.807	**0.758**	0.835	0.798	0.822
TB7(WN)	0.813	0.773	0.815	0.805	0.793	0.795	**0.432**	0.824	0.814
TB8(CA)	0.846	0.814	0.801	0.780	0.784	0.790	0.851	**0.763**	0.828
TB9(SE_2)	0.867	0.832	**0.696**	0.798	0.787	0.772	0.872	0.776	0.854



(a) 4SQDB (b) ACMDB

Figure 7: Histograms of s_b/s derived from brand conscious users learned by BITM.

1. $s = \frac{|U_{BC}|}{N_{users}}$ where U_{BC} is the set of all brand conscious users; s can be considered as the *average* ratio of brand conscious users.

2. For each brand b, we define the brand-specific ratio $s_b = \frac{|U^b_{BC}|}{N^b_{users}}$ where U^b_{BC} (N^b_{users}) is respectively the set of *brand conscious* users (the set of all users) who adopted items of brand b.

Note that on estimating s_b, we filtered out brands with $N^b_{users} < 5$ to avoid getting brands with large s_b by pure coincidence. After obtaining these quantities, we compare s_b of each brand with average value s using the ratio s_b/s.

Based on the distributions of the ratio s_b/s shown in Figure 7, we propose that exclusive brands (of both 4SQDB and ACMDB) are those for which $s_b/s \geq 2$. This is an appropriate threshold as an exclusive brand should have its s_b much higher than the average value s.

4SQDB: Recall that $N_{users} = 5406$ and BITM learned that $|U_{BC}| = 1319$, thus $s = 0.24$. There are 29 brands which satisfies $s_b/s \geq 2$. Thus, we can say that BITM learned 29 *exclusive* brands. In Table 8a, we show top-10 brands with largest s_b/s as representatives of exclusive brands for 4SQDB. We further checked the reliability of the result by looking at prices of these brands from sg.openrice.com, a popular website for rating food venues in Singapore. The prices are shown in the last column of Table 8a. Moreover, on comparing with another 29 less-exclusive brands (those with highest $s_b/s < 2$), the average price of exclusive brands is much higher than that of less-exclusive brands (**20.4** SGD compared with **9.8** SGD) while the standard deviation is comparable (**7.2** compared with **6.0**).

ACMDB: Recall that $N_{users} = 356$ and $|U_{BC}| = 58$, thus $s = 0.16$. Again, we determined exclusive authors as those whose $s_b/s \geq 2$. There are 23 authors satisfying this. Thus, BITM discovered 23 exclusive authors for this ACMDB dataset. Table 8b shows top-10 authors with largest s_b/s as representatives of exclusive authors. We further checked the reliability of the result by looking at h-index of these authors provided by Google Scholar. The h-indices are shown in the last column of Table 8b. Moreover, on comparing with another 23 *less-exclusive* authors (those with highest $s_b/s < 2$), the average h-index of exclusive authors is much higher than that of less-exclusive authors (**60.5** compared with **34.5**) while the standard deviation is smaller (**14.2** compared with **17.4**).

(a) 4SQDB

Brand	Area	s_b/s	Price (SGD)
The Halia	AB4	2.53	31-50
Ichiban Sushi	AB7	2.48	21-30
Sushi Tei	AB3	2.39	21-30
Nakhon Kitchen	AB8	2.29	11-20
ThaiExpress	AB12, AB10	2.19	11-20
Pepper Lunch	AB3	2.19	11-20
Pizza Hut	AB2	2.19	11-20
Sakae Sushi	AB8	2.19	11-20
Uncle Leong Seafood	AB5, AB6	2.19	11-20
Astons Specialities	AB9	2.18	11-20
Swensen's	AB1, AB11	2.18	11-20

(b) ACMDB

Author (i.e. brand)	**Topic**	s_b/s	**h-index**
Giovanni de Micheli	PO	5.00	73
Jon M. Kleinberg	WWW	4.06	69
David Karger	SE_1	4.03	70
Ion Stoica	WWW	3.98	65
Tian Zhang	DB + DM	3.87	59
Leslie Lamport	System	3.85	57
H. T. Kung	WN+ WWW	3.57	55
Jon Louis Bentley	SE_2	3.33	47
M. Frans Kaashoek	CA	3.13	45
John K. Ousterhout	SE_1	3.13	45

Table 8: Discovered exclusive brands for two datasets

5.5 Summary

Through the above analysis of topics and brand preferences, we demonstrate the usefulness of BITM model. Ideally, these empirical results should be further compared with ground truth topic and brand preference labels. In the absence of ground truth in 4SQDB and ACMDB, we further evaluate the BITM model in item adoption prediction task as described in Section 6.

6 Item Adoption Prediction Experiment

We define the item adoption prediction task as follows. For each user u with at least 4 item adoptions, we randomly hide p $(0 < p < 1)$ of these adoptions as the test data. The task is to predict these hidden item adoptions using the remaining $(1 - p)$ of adoptions to train a model.

Unlike in the standard recommendation problem where no item is rated again by the same user, the same item can be adopted by the same user in both training and test data. For example, the same paper can be cited by the same authors in multiple papers, and the same outlet can be checked-in multiple times by the same user.

We evaluate the prediction results using *average precision at* k ($AvgPrec@k$) which is defined to be the average of $Prec@k$ over all users with adoptions to be predicted. Let I_u^k be the top k predicted adopted items for user u ordered by $p(i_j|u)$, the probability of user i generating the adoption of item i_j. The precision at k for user u, $Prec@k(u)$, is defined as:

$$Prec@k(u) = \frac{|Test_u \cap I_u^k|}{k}$$

where $Test_u$ denotes the set of item adoptions of user u to be predicted.

To ensure the results are robust, we conducted 4-fold and 5-fold cross validation of the training and testing data for 4SQDB and ACMDB respectively, and reported the average results. We vary k from 2 to 2000 for 4SQDB, and from 1 to 3000 for ACMDB.

Other than BITM and LDA, we also introduce two other simple baselines, namely:

- *Global Popularity* (**GPOP**): Each item is assigned a global popularity score defined by the number of adoptions it has. Usually, GPOP is not appropriate for prediction task that involves items of very different characteristics. In this experiment, however, the items involved are similar. We therefore include GPOP and also include the local popularity score below.
- *Local Popularity* (**LPOP**): For each user, we assign each item a local popularity score defined by the number of adoptions the user has performed on the item. The items are then ranked by decreasing local popularity score. For each user, his top ranked items are returned as the predicted adoptions.

Note that GPOP returns the same adoption predictions for all users while LPOP returns the frequently adopted items by the target user.

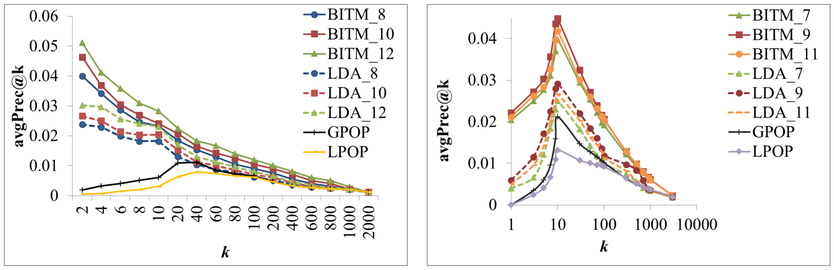
Figures 8a and 8b show the $AvgPrec@k$ of prediction results on 4SQDB and ACMDB respectively by varying k and number of topics. The key observations from these figures are that:

- BITM is consistently the best performing model and it is followed by LDA, GPOP and LPOP. We observe this for both datasets for almost all k's and all number of topics.
- For **4SQDB**, the $AvgPrec@k$ of BITM and LDA decreases with increasing k. This suggests that the top ranked predicted adoptions by the two models are more accurate than the lower ranked predicted adoptions. For **ACMDB**, we however observe that $AvgPrec@k$ increases initially until k reaches about 10. Beyond that, $AvgPrec@k$ decreases with larger k. This observation holds for all the models.
- The optimal number of topics for both BITM and LDA for the **4SQDB** dataset is 12 while that for **ACMDB** is 9. This observation is consistent with the numbers of topics determined for the two datasets by likelihood.

To sum up, BITM shows promising prediction results in this experiment and the results are also consistent for both datasets under across different settings.

7 Conclusion

Brand-based item adoption decisions are important knowledge that has widespread applications in marketing and ad-



(a) **4SQDB** (b) **ACMDB**

Figure 8: Prediction results w.r.t different number of topics for two datasets. $BITM_i$ (LDA_i) are $BITM$ (LDA) trained with i topics respectively.

vertising. In this paper, we introduce BITM model as a novel way to learn the hidden user-specific, item specific and brand specific variables that explain an observed set of item adoptions. BITM does not require price information in training data. Using both synthetic and real datasets, we show that BITM can effectively learn the hidden variables recovering the exclusive brands and brand conscious users, and can be used to predict item adoptions more accurately than the baseline model. As BITM is the first model that considers brand effects, one can pursue further research based on this work. In particular, we can extend BITM to consider constraints that allow the model to learn different kinds of exclusive brands such as those that are expensive in price, or those that are adopted by users of some categories. This will give rise to a range of different models that can be applied to different adoption scenarios. Finally, further analysis on BITM results can also reveal customers' loyalty to any interested brand, which is extremely important for planning future marketing strategies.

Acknowledgement

This research is supported by the Singapore National Research Foundation under its International Research Centre @ Singapore Funding Initiative and administered by the IDM Programme Office.

References

Aaker, D. A. 2012. Brand extensions: the good, the bad and the ugly. *Sloan Management Review* 31(4).

2011. ACM Digital Library. http://dl.acm.org/.

Blei, D. M.; Ng, A. Y.; and Jordan, M. I. 2003. Latent dirichlet allocation. *J. Mach. Learn. Res.* 3.

Casella, G., and George, E. I. 1992. Explaining the Gibbs sampler. *The American Statistician* 46(3):167–174.

Chaudhuri, A., and Holbrook, M. B. 2001. The chain of effects from brand trust and brand affect to brand performance: the role of brand loyalty. *The Journal of Marketing* 65(2):81–93.

Dempster, A. P.; Laird, N. M.; and Rubin, D. B. 1977. Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)* 39(1):1–38.

Dick, A. S., and Basu, K. 1994. Customer loyalty: toward an integrated conceptual framework. *Journal of the Academy of Marketing Science* 22(2):99–113.

Erdem, T., and Keane, M. P. 1996. Decision-making under uncertainty: Capturing dynamic brand choice processes in turbulent consumer goods markets. *Marketing Science* 15(1):pp. 1–20.

Erdem, T.; Swait, J.; Broniarczyk, S.; Chakravarti, D.; Kapferer, J.-N.; Keane, M.; Roberts, J.; Steenkamp, J.-B. E.; and Zettelmeyer, F. 1999. Brand equity, consumer learning and choice. *Marketing Letters* 10(3):301–318.

Gelfand, A. E., and Smith, A. F. 1990. Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association* 85(410).

Heinrich, G. 2005. Parameter estimation for text analysis. http://www.arbylon.net/publications/text-est.pdf.

Hoffman, M.; Bach, F. R.; and Blei, D. M. 2010. Online learning for latent dirichlet allocation. In *NIPS*.

Hofmann, T. 2001. Learning what people (don't) want. In *ECML*.

Hofmann, T. 2003. Collaborative filtering via gaussian probabilistic latent semantic analysis. In *SIGIR*.

Hofmann, T. 2004. Latent semantic models for collaborative filtering. *ACM Trans. Inf. Syst.* 22(1):89–115.

Jacoby, J., and Chestnut, R. W. 1978. *Brand loyalty: Measurement and management*. Wiley New York.

Jin, R.; Si, L.; and Zhai, C. 2002. Preference-based graphic models for collaborative filtering. In *UAI*.

Liu, N. N.; Zhao, M.; and Yang, Q. 2009. Probabilistic latent preference analysis for collaborative filtering. In *CIKM*.

Luu, D. M. 2013a. Derivations of Gibbs sampling for BITM. http://goo.gl/8WcEGY.

Luu, D. M. 2013b. Topics and their corresponding top papers learned by BITM. http://goo.gl/nuMnIw.

Redner, R. A., and Walker, H. F. 1984. Mixture densities, maximum likelihood and the em algorithm. *SIAM review* 26(2):195–239.

Smith, M. D., and Brynjolfsson, E. 2001. Consumer decision-making at an internet shopbot: Brand still matters. *The Journal of Industrial Economics* 49(4):541–558.