Data Min Knowl Disc
DOI 10.1007/s10618-012-0281-y

Topic model for analyzing purchase data with price information

Tomoharu Iwata · Hiroshi Sawada

Received: 17 November 2011 / Accepted: 13 June 2012

Abstract We propose a new topic model for analyzing purchase data with price information. Price is an important factor in consumer purchase behavior. The proposed model assumes that a topic has its own price distributions for each item as well as an item distribution. The topic proportions, which represent a user's purchase tendency, are influenced by the user's purchased items and their prices. By estimating the mean and the variance of the price for each topic, the proposed model can cluster related items taking their price ranges into consideration. We present its efficient inference procedure based on collapsed Gibbs sampling. Experiments on real purchase data demonstrate the effectiveness of the proposed model.

Keywords Purchase log · Probabilistic topic modeling · Gibbs sampling · Clustering

1 Introduction

It is important to model user purchase behavior because it helps us to find a marketing strategy and perform trend analysis, and provides recommendations and personalized advertisements. Topic models have been used for analyzing log data of human behavior such as web usage (Jin et al. 2004) and news article reading (Das et al. 2007) as well as purchase behavior (Iwata et al. 2009). A topic model is a probabilistic generative model for discrete data, where each user has his/her own probability distribution of selecting each topic, and each topic has a purchase probability distribution over

Responsible editor: Bing Liu.

T. Iwata (✉) · H. Sawada
NTT Communication Science Laboratories, 2-4, Hikaridai, Seika-cho, Soraku-gun,
Kyoto 619-0237, Japan
e-mail: iwata.tomoharu@lab.ntt.co.jp

Published online: 27 June 2012

items. Topic models are mainly used for analyzing text data, where users and items in purchase data correspond to documents and words in text data, respectively. "Topic" in purchase data represents a latent category of items that are likely to be purchased by a particular set of users. By using topic models, we can obtain a low dimensional representation for each user, and find latent purchase patterns of co-occurrence items.

In this paper, we propose a new topic model for analyzing purchase data with price information. Price is an important factor in consumer purchase behavior. People sometimes give up the idea of buying favorite items because they are too expensive, and buy reasonably priced items even if they do not greatly care for them. The proposed model has a distribution over items for each topic, in the same way as standard topic models, such as latent Dirichlet allocation (LDA) (Blei et al. 2003). In addition, the proposed model assumes that each topic has its own price distribution for each item, where the distributions have different means and variances depending on topics and items. Thus, topic proportions, which represent a user's purchase tendency, are influenced by the items purchased by the user and their prices.

Concurrently purchased items can differ depending on their prices. For example, high ranking wines might often be purchased with expensive cheeses and hams; on the other hand, cheap wines might be purchased with vegetables for use in cooking. The price range can also affect purchase behaviors in relation to other items. For example, fashionable people who buy clothes in a wide price range might often buy fashion magazines; thrifty people who buy only cheap clothes are not likely to buy fashion magazines. By inferring price distributions with their means and variances as well as item distributions, the proposed model can find clusters of concurrently purchased items taking their price ranges into consideration.

The proposed model is a Bayesian probabilistic model. We use conjugate priors for all of the parameters in the proposed model, i.e. Dirichlet priors for multinomial parameters, and Gaussian-Gamma priors for Gaussian parameters. Therefore, the inference can be efficiently performed based on collapsed Gibbs sampling by integrating out the parameters.

Standard topic models, such as LDA, are used for modeling samples (e.g. users and documents), each of which is represented by a set of discrete values (e.g. purchased items and words). More recently, topic models have been proposed that incorporate other information such as authors (Rosen-Zvi et al. 2004), time (Wang and McCallum 2006; Blei and Lafferty 2006; Iwata et al. 2010) and annotations (Blei and Jordan 2003). The proposed model is different from these models in that it can model sets of items where a continuous value (price) is associated with each item. Standard topic models can cluster items that are likely to be purchased by a specific user group. However, they do not distinguish differences in price. One approach can model the price distribution of an item by using a Gaussian mixture, and it can find clusters of price ranges. However, it does not cluster concurrently purchased items whose purchase is price dependent. The proposed model resolves these problems by incorporating price distributions into topic models and inferring item and price distributions simultaneously.

The remainder of this paper is organized as follows. In Sect. 2, we propose a topic model for purchase data with price information. In Sect. 3, we present its efficient inference procedure based on collapsed Gibbs sampling. In Sect. 4, we outline related work. In Sect. 5, we demonstrate the effectiveness of the proposed method by using real purchase data. Finally, we present concluding remarks and a discussion of future work in Sect. 6.

2 Proposed model

Suppose that we have a set of purchase logs of U users. The purchase log of user u consists of purchased items and their prices $(\boldsymbol{x}_u, \boldsymbol{v}_u)$. Here, $\boldsymbol{x}_u = \{x_{un}\}_{n=1}^{N_u}$ represents items that are purchased by user u, and $\boldsymbol{v}_u = \{v_{un}\}_{n=1}^{N_u}$ represents their prices. Our notation is summarized in Table 1.

The proposed model finds latent topics that are influenced by both concurrently purchased items and their prices. The proposed model first generates an item to be purchased, and then generates its price. The generative process for purchased items is the same as that of standard topic models. Each user has topic proportions $\boldsymbol{\theta}_u$ that are sampled from a Dirichlet distribution. For each of the N_u purchases, a topic z_{un} is chosen from the topic proportions, and then item x_{un} is purchased according to a topic-specific multinomial distribution $\boldsymbol{\phi}_{z_{un}}$.

Each topic has its own price distribution for each item, which is assumed to be a Gaussian distribution, $\text{Normal}(\mu_{ki}, \lambda_{ki}^{-1})$. By setting different means and precisions depending on the topics, we can analyze the price and its range that are specific to the topic. The mean μ_{ki} and the precision λ_{ki} are sampled respectively from Gaussian and Gamma distributions, which are conjugate priors of a Gaussian distribution. After topic z_{un} and item x_{un} are sampled, its price v_{un} is determined according to the topic-specific Gaussian distribution, $\text{Normal}(\mu_{z_{un}x_{un}}, \lambda_{z_{un}x_{un}}^{-1})$.

In summary, the proposed model assumes the following generative process for a set of purchase logs with price information $\boldsymbol{X} = \{\boldsymbol{x}_u\}_{u=1}^{U}$ and $\boldsymbol{V} = \{\boldsymbol{v}_u\}_{u=1}^{U}$:

1. For each topic $k = 1, \ldots, K$:
   (a) Draw item probability $\boldsymbol{\phi}_k \sim \text{Dirichlet}(\beta)$
   (b) For each item $i = 1, \ldots, I$:

Table 1 Notation

Symbol	Description
U	Number of users
I	Number of items
K	Number of latent topics
N_u	Number of purchased items of user u
x_{un}	nth purchased item of user u
v_{un}	Price of nth purchased item of user u
z_{un}	Latent topic of nth purchased item of user u
$\boldsymbol{\theta}_u$	Topic proportions for user u, $\boldsymbol{\theta}_u = \{\theta_{uk}\}_{k=1}^{K}, \theta_{uk} \geq 0, \sum_k \theta_{uk} = 1$
$\boldsymbol{\phi}_k$	Multinomial distribution over items for topic k, $\boldsymbol{\phi}_k = \{\phi_{ki}\}_{i=1}^{I}, \phi_{ki} \geq 0, \sum_i \phi_{ki} = 1$
μ_{ki}	Mean price of item i for topic k
λ_{ki}	Precision (inverse of variance) of price of item i for topic k

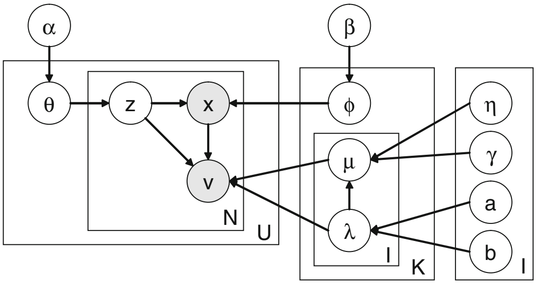

Fig. 1 Graphical model representation of the proposed topic model for purchase data with price information

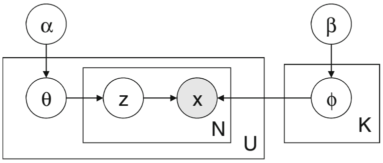

Fig. 2 Graphical model representation of the standard topic model

i. Draw price precision $\lambda_{ki} \sim \text{Gamma}(a_i, b_i)$
ii. Draw price mean $\mu_{ki} \sim \text{Normal}(\eta_i, (\gamma_i \lambda_{ki})^{-1})$

2. For each user $u = 1, \ldots, U$:
   (a) Draw topic proportions $\boldsymbol{\theta}_u \sim \text{Dirichlet}(\alpha)$
   (b) For each transaction $n = 1, \ldots, N_u$:
      i. Draw topic $z_{un} \sim \text{Multinomial}(\boldsymbol{\theta}_u)$
      ii. Draw item $x_{un} \sim \text{Multinomial}(\boldsymbol{\phi}_{z_{un}})$
      iii. Draw price $v_{un} \sim \text{Normal}(\mu_{z_{un}x_{un}}, \lambda^{-1}_{z_{un}x_{un}})$,

where α and β are Dirichlet parameters, $\boldsymbol{a} = \{a_i\}_{i=1}^{I}$ and $\boldsymbol{b} = \{b_i\}_{i=1}^{I}$ are shape and inverse scale parameters of Gamma distributions, respectively, and $\boldsymbol{\eta} = \{\eta_i\}_{i=1}^{I}$ and $\boldsymbol{\gamma} = \{\gamma_i\}_{i=1}^{I}$ are mean and precision parameters of Gaussian distributions, respectively.

Figure 1 shows a graphical model representation of the proposed topic model for purchase data, which represents dependencies among variables. Here the shaded and unshaded nodes indicate observed and latent variables, respectively, the plate indicates replicates, and the value in the plate indicates the number of replicates. For comparison, we present a graphical model of the standard topic model, or latent Dirichlet allocation in Fig. 2. The proposed model is an extension of the standard topic model, where price v is generated depending on topic z and item x. In addition, priors for parameters of Gaussian distribution are also incorporated for robust Bayesian inference.

The joint distribution of purchased items $\boldsymbol{X}$, their prices $\boldsymbol{V}$, and latent topics $\boldsymbol{Z} = \{\boldsymbol{z}_u\}_{u=1}^{U}$, where $\boldsymbol{z}_u = \{z_{un}\}_{n=1}^{N_u}$, is described as follows:

$$P(\boldsymbol{X}, \boldsymbol{V}, \boldsymbol{Z}|\alpha, \beta, \boldsymbol{\eta}, \boldsymbol{\gamma}, \boldsymbol{a}, \boldsymbol{b}) = P(\boldsymbol{Z}|\alpha)P(\boldsymbol{X}|\boldsymbol{Z}, \beta)P(\boldsymbol{V}|\boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{\eta}, \boldsymbol{\gamma}, \boldsymbol{a}, \boldsymbol{b}). \quad (1)$$

We can integrate out multinomial parameters in the first and second factors, $\{\boldsymbol{\theta}_u\}_{u=1}^{U}$ and $\{\boldsymbol{\phi}_k\}_{k=1}^{K}$, because we use Dirichlet distributions for their priors, which are conjugate to multinomial distributions. The first factor on the right hand side of (1) is calculated by:

$$P(\boldsymbol{Z}|\alpha) = \prod_{u=1}^{U} \int \prod_{n=1}^{N_u} P(z_{un}|\boldsymbol{\theta}_u) P(\boldsymbol{\theta}_u|\alpha) d\boldsymbol{\theta}_u, \tag{2}$$

and we have the following equation by integrating out $\{\boldsymbol{\theta}_u\}_{u=1}^{U}$:

$$P(\boldsymbol{Z}|\alpha) = \left(\frac{\Gamma(\alpha K)}{\Gamma(\alpha)^K}\right)^U \prod_u \frac{\prod_k \Gamma(N_{ku} + \alpha)}{\Gamma(N_u + \alpha K)}, \tag{3}$$

where $\Gamma(\cdot)$ is the gamma function, and N_{ku} is the number of purchased items assigned to topic k in the history of user u. Similarly, the second factor is given as follows by integrating out $\{\boldsymbol{\phi}_k\}_{k=1}^{K}$:

$$P(\boldsymbol{X}|\boldsymbol{Z}, \beta) = \left(\frac{\Gamma(\beta I)}{\Gamma(\beta)^I}\right)^K \prod_k \frac{\prod_i \Gamma(N_{ki} + \beta)}{\Gamma(N_k + \beta I)}, \tag{4}$$

where N_{ki} is the number of times item i has been assigned to topic k, and $N_k = \sum_i N_{ki}$.

We can also integrate out Gaussian parameters in the third factor of (1), $\{\{\mu_{ki}\}_{i=1}^{I}\}_{k=1}^{K}$ and $\{\{\lambda_{ki}\}_{i=1}^{I}\}_{k=1}^{K}$, because we use Gaussian and Gamma distributions for their priors, which are conjugate to Gaussian distributions. The third factor is given as follows:

$$P(\boldsymbol{V}|\boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{\eta}, \boldsymbol{\gamma}, \boldsymbol{a}, \boldsymbol{b}) = \prod_{k=1}^{K} \prod_{i=1}^{I} P(\boldsymbol{V}_{ki}|\boldsymbol{\eta}, \boldsymbol{\gamma}, \boldsymbol{a}, \boldsymbol{b}), \tag{5}$$

where $\boldsymbol{V}_{ki}$ is a set of prices of item i that is assigned to topic k, and

$$\begin{aligned} &P(\boldsymbol{V}_{ki}|\boldsymbol{\eta}, \boldsymbol{\gamma}, \boldsymbol{a}, \boldsymbol{b}) \\ &= \int\int \prod_{v \in \boldsymbol{V}_{ki}} P(v|\mu_{ki}, \lambda_{ki}^{-1}) P(\mu_{ki}|\eta_i, (\gamma_i \lambda_{ki})^{-1}) P(\lambda_{ki}|a_i, b_i) d\mu_{ki} d\lambda_{ki} \\ &= (2\pi)^{-\frac{N_{ki}}{2}} \frac{\Gamma(a_{ki})}{\Gamma(a_i)} \frac{b_i^{a_i}}{b_{ki}^{a_{ki}}} \left(\frac{\gamma_i}{\gamma_{ki}}\right)^{\frac{1}{2}}. \end{aligned} \tag{6}$$

Here, η_{ki} and γ_{ki} are hyperparameters of posterior distributions for mean μ_{ki}, $P(\mu_{ki}|\boldsymbol{X}, \boldsymbol{V}, \boldsymbol{Z}, \eta_i, \gamma_i) = \text{Normal}(\eta_{ki}, \gamma_{ki}^{-1})$, and a_{ki} and b_{ki} are hyperparameters of posterior distributions for precision λ_{ki}, $P(\lambda_{ki}|\boldsymbol{X}, \boldsymbol{V}, \boldsymbol{Z}, a_i, b_i) = \text{iGamma}(a_{ki}, b_{ki})$. They are given as follows:

$$\eta_{ki} = \frac{\gamma_i \eta_i + \sum_{v \in \boldsymbol{V}_{ki}} v}{\gamma_i + N_{ki}}, \tag{7}$$

$$\gamma_{ki} = \gamma_i + N_{ki}, \tag{8}$$

$$a_{ki} = a_i + \frac{N_{ki}}{2}, \tag{9}$$

$$b_{ki} = b_i + \frac{\sum_{v \in \boldsymbol{V}_{ki}} v^2}{2} + \frac{\gamma_i \eta_i^2}{2} - \frac{\gamma_{ki} \eta_{ki}^2}{2}. \tag{10}$$

In this way, the parameters in the proposed model can be integrated out, and thus, an efficient inference procedure can be derived as described in the next section.

3 Inference by Gibbs sampling

Given a set of purchased items with their prices, $\boldsymbol{X}$ and $\boldsymbol{V}$, we would like to infer the latent topics $\boldsymbol{Z}$. Several approaches have been proposed for inferring topic models, such as variational Bayes (Blei et al. 2003), collapsed Gibbs sampling (Griffiths and Steyvers 2004), expectation propagation (Minka and Lafferty 2002), and collapsed variational Bayes (Teh et al. 2006b). In this section, we present a simple and efficient inference procedure based on collapsed Gibbs sampling.

Given the current state of all but one variable z_j, where $j = (u, n)$, the assignment of a latent topic to the nth purchased item of user u is sampled from the following probability:

$$\begin{aligned} &P(z_j = k | \boldsymbol{X}, \boldsymbol{V}, \boldsymbol{Z}_{\setminus j}, \alpha, \beta, \boldsymbol{\eta}, \boldsymbol{\gamma}, \boldsymbol{a}, \boldsymbol{b}) \\ &\propto \frac{N_{ku \setminus j} + \alpha}{N_{u \setminus j} + \alpha K} \frac{N_{k x_j \setminus j} + \beta}{N_{k \setminus j} + \beta I} \frac{\Gamma(a_{k x_j})}{\Gamma(a_{k x_j \setminus j})} \frac{b_{k x_j \setminus j}^{a_{k x_j \setminus j}}}{b_{k x_j}^{a_{k x_j}}} \left(\frac{\gamma_{k x_j \setminus j}}{\gamma_{k x_j}} \right)^{\frac{1}{2}}, \end{aligned} \tag{11}$$

where $\setminus j$ represents the count or hyperparameter when excluding sample j. See Appendix A for the derivation. The first and second factors in (11) are the same as those of LDA. The remaining factors are derived from price distributions. The hyperparameters of Gaussian-Gamma distributions excluding sample j are calculated as follows:

$$\eta_{ki \setminus j} = \frac{\gamma_{ki} \eta_{ki} - v_j}{\gamma_{ki} - 1}, \tag{12}$$

$$\gamma_{ki \setminus j} = \gamma_{ki} - 1, \tag{13}$$

$$a_{ki \setminus j} = a_{ki} - \frac{1}{2}, \tag{14}$$

$$b_{ki \setminus j} = b_{ki} - \frac{v_j^2}{2} + \frac{\gamma_{ki} \eta_{ki}^2}{2} - \frac{\gamma_{ki \setminus j} \eta_{ki \setminus j}^2}{2}. \tag{15}$$

The hyperparameters of Dirichlet distributions, α and β, can be estimated by maximizing the joint likelihood (1) by using the fixed-point iteration method described in (Minka 2000) as follows:

$$\alpha \leftarrow \alpha \frac{\sum_u \sum_k \Psi(N_{ku} + \alpha) - UK\Psi(\alpha)}{K\left(\sum_u \Psi(N_u + \alpha K) - U\Psi(\alpha K)\right)}, \tag{16}$$

$$\beta \leftarrow \beta \frac{\sum_k \sum_i \Psi(N_{ki} + \beta) - KI\Psi(\beta)}{I\left(\sum_k \Psi(N_k + \beta I) - K\Psi(\beta I)\right)}, \tag{17}$$

where $\Psi(\cdot)$ is the digamma function defined by $\Psi(x) = \frac{\partial \log \Gamma(x)}{\partial x}$. We can set the hyperparameters of Gaussian-Gamma distributions, $\{\eta_i, \gamma, a_i, b_i\}_{i=1}^{I}$, by using the empirical means and variances of the given data. In the experiments, we used $\eta_i = \bar{\mu}_i, \gamma_i = 1, a_i = 1$, and $b_i = 1 + \bar{\sigma}_i^2$, where $\bar{\mu}_i$ and $\bar{\sigma}_i^2$ are the empirical mean and variance of the price of item i.

At the beginning of the inference, we initialize latent topics so that none of the purchased items are assigned to any topics. In the first iteration, a latent topic is sampled according to (11) for each item, where it is unnecessary to exclude the current sample because it has no topic. After that, by iterating Gibbs sampling with (11) and maximum likelihood estimation with (16) and (17), we can infer latent topics while optimizing the parameters.

We can estimate parameters $\theta_{uk}, \phi_{ki}, \mu_{ki}$ and λ_{ki} using sampled latent topics $\boldsymbol{Z}$ as follows:

$$\hat{\theta}_{uk} = \frac{N_{ku} + \alpha}{N_u + \alpha K}, \tag{18}$$

$$\hat{\phi}_{ki} = \frac{N_{ki} + \beta}{N_k + \beta I}, \tag{19}$$

$$\hat{\mu}_{ki} = \eta_{ki}, \tag{20}$$

$$\hat{\lambda}_{ki} = \frac{a_{ki}}{b_{ki}}. \tag{21}$$

The estimated topic proportions $\{\hat{\theta}_{uk}\}_{k=1}^{K}$ can be used for a low dimensional representation of user u, and the estimated parameters $\{\hat{\phi}_{ki}, \hat{\mu}_{ki}, \hat{\lambda}_{ki}\}_{i=1}^{I}$ can be used to analyze the characteristics of topic k.

4 Related Work

The proposed model is latent Dirichlet allocation (LDA) (Blei et al. 2003) extended to include price information. Topic models are successfully used for a wide variety of applications including information retrieval (Blei et al. 2003; Hofmann 1999), collaborative filtering (Hofmann 2003), language modeling (Wallach 2006; Williamson et al. 2010), network analysis (Fu et al. 2009) and multilingual text analysis (Mimno 2009) as well as for modeling human behavior (Jin et al. 2004; Das et al. 2007; Iwata et al. 2009). A number of topic models have been proposed for incorporating different types of information, such as author topic model (ATM) (Rosen-Zvi et al. 2004), correspondence latent Dirichlet allocation (corr-LDA) (Blei and Jordan 2003), and topic over time (TOT) (Wang and McCallum 2006). ATM and corr-LDA integrate authors and annotations into topic models, respectively. The authors and annotations are discrete variables, and they are associated with each document, but not each word. In this paper, we analyze purchase data, in which a continuous variable (price) is associated with each item (word), and prices differ among purchased items of a user (document). TOT is a topic model for incorporating a continuous variable. TOT captures the dynamics of topics by assuming that each topic has its own beta distribution that generates the document's time stamp. In TOT, the time stamp is assumed to be associated with each word, although it is usually constant across the document. And the time stamp is generated depending only on the topic; time is conditionally independent of words given the topic. On the other hand, the proposed model assumes that the price is generated depending on both the item (word) and the topic, which is natural for a generative process of prices. There have been some work on analyzing point of sales (POS) data, which focus mainly on finding frequent purchase patterns and discovering general sales trends (Hamuro et al. 1998; Thomas and Chakravarthy 2000). Modeling the effect of price promotion, which is studied in marketing science (Raju 1992; Nijs et al. 2001), is also related to our work. However, the aim is different from ours, which is to model purchase behavior for each user. Purchase data are often used to enable recommendation methods to learn preference, or purchase behavior, for each user (Shani et al. 2005; Iwata et al. 2006; Li et al. 2009). However, these methods do not take price information into account.

5 Experimental Results

We evaluated the proposed topic model using real purchase data with price information. The data consist of records in an online service for managing household accounts from February 2009 to November 2010 in Japan. The user can keep his/her household accounts by recording the purchased items with their prices in the online service. The data were noisy; since the users can freely choose the name of an item, an item can be referred to by different names among users, and different items can be referred to by the same name. For simplicity, we assumed that if the names were different, the items were different. To protect privacy, we omitted items that were purchased by fewer than 300 users. We also omitted users who purchased fewer than 30 items. The data consisted of 5, 052 items, and 48, 427 users.

Table 2 is the topic extraction result, which shows ten items that are the most likely to be purchased in each topic. They are estimated with the proposed model using 100 topics. We translated the Japanese names into English. The right and left values in each column represent the mean price in yen and its standard deviation for the item in the topic. We can see that related items are clustered, for example, Topic1 is about drink, Topic2 is about baby supplies, and Topic3 is about food. For these topics, items in the same product category are clustered. Some other topics represent lifestyle, and items from different categories are clustered. For example, Topic5 contains stockings, cosmetics and magazines, which fashion conscious women are likely to buy. Topic6 includes beer, pachinko and cigarettes, which are items men who like gambling might buy. Topic11 contains drinking party, taxi and cleaning, which are items company workers are likely to buy. From the topic extraction result, we can also see that there are some topic specific price ranges. Certain items have different price ranges depending on topics. For example, the price of lunch in Topic5 is 1,189 yen, and that in Topic7 is 757 yen; this indicates that fashion conscious women spend more money on lunch in fancy restaurants. In this purchase data, items with the same name might indicate different things, because users can freely choose the name of an item. We can distinguish them from the result. For example, curry and ramen (noodles) in Topic4 indicate that users have eaten them in restaurants; those in Topic9 indicate that users

Table 2 Some examples of extracted topics obtained with the proposed method

Topic 1			Topic 2			Topic 3		
Canned coffee	103	17	Nappy	1181	253	Sprout	35	7
Vegetable juice	90	18	Diaper	1158	235	Natto	93	19
Cocktail	173	25	Parts	293	124	Tofu	98	31
canned beer	207	97	Baby wipe	416	216	Cucumber	54	13
Dietary supplement	123	53	Weaning food	152	79	Enokidake	71	21
Canned juice	127	54	Latte	306	148	Onion	104	51
Tea	96	8	Daiso	123	80	Carrot	127	45
Ice	165	77	Baby food	213	122	Chivee	96	7
Yogurt	46	19	Milk	1462	533	Kimchi	251	78
Sweet bun	206	62	Sweet bun	179	48	Shiitake	157	45
Topic 4			Topic 5			Topic 6		
Box lunch	416	128	Stocking	646	278	Beer	734	334
Ramen	726	185	Box lunch	469	142	Pachinko	17864	15353
Curry	627	295	Magazine	624	130	Cigarette	790	289
Wheat noodle	473	154	**Lunch**	1189	363	Juice	365	156
Consumables	320	385	Beauty wash	1615	1037	Sushi	1965	1310
Green tee	154	101	Pasta	961	416	Box lunch	747	254
Beef-on-rice	394	122	Chocolate	182	115	Low-malt beer	611	214
Noodle	539	266	Eyelash liner	1307	363	Lottery	2013	1712
Coffee	333	154	Napkin	244	97	Grilled beef	6941	3350
Pasta	453	208	Tights	823	443	Prepared food	1874	1138
Topic 7			Topic 8			Topic 9		
Lunch	757	352	Coffee	104	27	Natto	76	17
Coffee	303	132	Green tea	119	31	Wheat noodle	69	56
Magazine	360	165	Tea	116	51	Tofu	66	16
Tea	456	255	Ice coffee	103	17	Ice	172	60
Dinner	1884	1532	Vegetable juice	178	58	**Ramen**	174	96
Food shopping	1208	941	Sports drink	134	40	Toilet tissue	262	65
Subway	310	267	Bread	177	35	**Curry**	128	73
Snack	599	209	Calpis	150	60	Pasta	178	92
Haircut	5722	2996	Snack	117	27	Toothbrush	110	47
Taxi	2508	1090	Ice-cream	316	141	Chocolate	169	88
Topic 10			Topic 11			Topic 12		
Socks	689	367	Drink party	3499	1552	Pudding	100	29
Bread	218	92	Taxi	918	397	Ice	215	70
Pants	1134	1028	Shirt cleaning	141	146	Cream puff	99	27
T-shirts	908	530	Rental DVD	218	186	Rice cracker	145	86
Juice	227	108	Drink money	6016	3490	Cookie	135	67

Table 2 Continued

Topic 10			Topic 11			Topic 12		
One-piece	2293	1783	Golf	9266	5305	Sushi	495	450
Sandal	1557	1160	Cleaning	902	625	Sweet bun	104	31
T-shirts	1058	645	Bullet train	7492	4014	Snack	97	15
Movie	1173	360	Mobile charge	5332	2885	Cheese	203	78
Legging	697	342	Water charge	2465	1106	Chips	97	23

Each *column* represents a topic, which shows ten items that are most likely to be purchased. The *right* and *left* values in each *column* represent the mean price in yen and the standard deviation of the item in the topic. The same item names with different price ranges are in *bold*

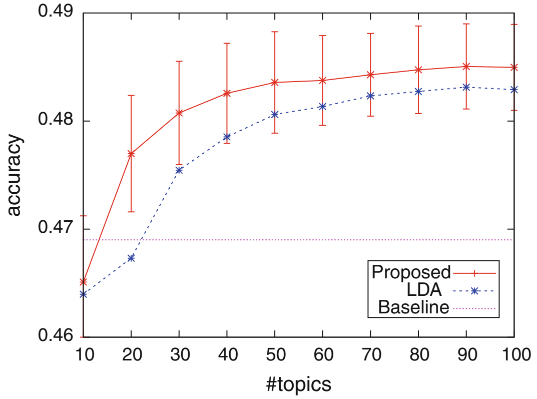

Fig. 3 Accuracies of predicting the age when using topic proportions as input

have bought uncooked curry and ramen in grocery stores. We can determine this from the price range and other items in the topic. In these ways, the proposed model can cluster items taking their price ranges into consideration.

For a quantitative evaluation measure, we calculated the accuracy of predicting the user's age when using topic proportions as input. The ages of some users were provided. Age was divided into sections, where the width of each section was ten years. We predicted the age section of users using k-nearest neighbor classifiers. For finding neighbors, we used Kullback–Leibler (KL) divergence between topic proportions that were estimated with the proposed model. In particular, neighbors u' of user u were selected by minimizing the following equation, $\sum_{k=1}^{K} \theta_{uk} \log \frac{\theta_{uk}}{\theta_{u'k}}$. The high prediction accuracy implies that the estimated topic proportions appropriately represent intrinsic characteristics of the user. We compared the proposed model and LDA by using their topic proportions as input. We generated 100 sets of training data, each of which contained the purchase logs of 3,000 users that were randomly sampled. Figure 3 shows the results obtained with different numbers of topics, where we set the number of neighbors at $k = 100$. The baseline plot represents the accuracies when the set of purchased items was used as the input of the k-nearest neighbor classifiers. The proposed model did not achieve the significant improvement over the LDA; the standard deviation bars of the proposed model overlapped with the accuracies of

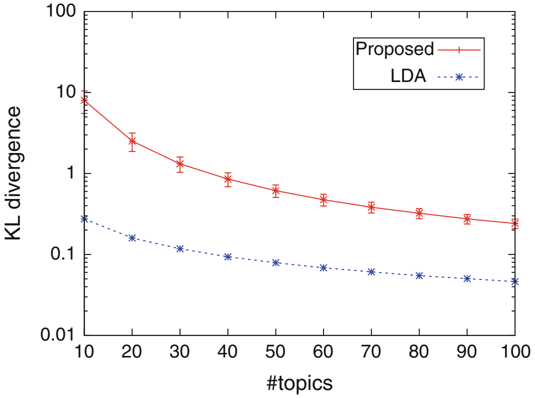

Fig. 4 KL divergences of price distributions between different topics. The x-axis is the number of topics

LDA. However, the proposed model outperformed LDA for all numbers of topics, and achieved significant improvement over the baseline method. This result indicates that the proposed model can learn information about users adequately by incorporating price information into topic models.

We might obtain the topic specific price ranges by calculating that means and standard deviations after extracting topics using LDA. We compared this approach with the proposed model in terms of specificity of price ranges for each topic. The specificity is evaluated by using the following average KL divergence of price distributions between different topics,

$$\frac{1}{IK(K-1)}\sum_{i=1}^{I}\sum_{k=1}^{K}\sum_{k'\neq k}\int P(v|\mu_{ki},\lambda_{ki}^{-1})\log\frac{P(v|\mu_{ki},\lambda_{ki}^{-1})}{P(v|\mu_{k'i},\lambda_{k'i}^{-1})}dv, \tag{22}$$

which is averaged over all items and all pairs of topics. The specificity becomes high when the price distributions differ depending on topics. Figure 4 shows the result. The KL divergences of the proposed model were higher than those of LDA for all numbers of topics. This result implies that we can extract characteristic topics that have specific price ranges by inferring topics using both item and price information.

Figure 5 shows the log joint likelihoods (1) over iterations in the inference with different numbers of topics ($K = 10, 30, 50, 100$). As the number of iterations increases, the log likelihood increases, and eventually converges to a certain point, although there are small fluctuations because the inference is based on Gibbs sampling. This result indicates that the inference procedures described in Section 3 can appropriately find latent topics and parameters that fit with the given purchase log data. The log likelihoods drastically increases in the small number of iterations.

6 Conclusion

We have proposed a topic model for analyzing purchase data with price information and its efficient inference procedure. We have confirmed experimentally that the

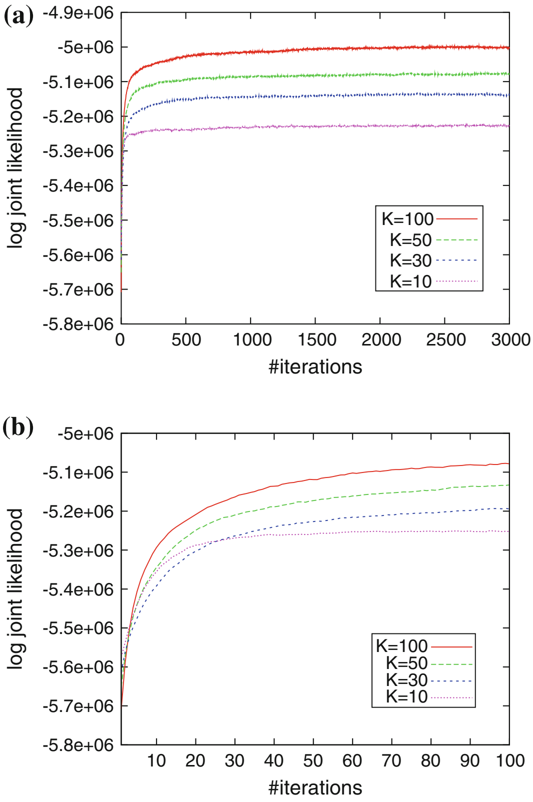

Fig. 5 Log joint likelihoods over iterations with different numbers of topics (**a**), and its zoom of 0 to 100 iterations (**b**)

proposed model can learn intrinsic user characteristics, and can cluster related items with their prices.

Although our results have been encouraging to date, our model can be further improved in a number of ways. First, we can automatically infer the number of topics by extending the model to a nonparametric Bayesian model such as hierarchical Dirichlet processes (Teh et al. 2006a). Second, we can incorporate information other than price into topic models to allow us to analyze purchase data. Purchase data might contain other information, such as purchase times, stores, types of payments and user attributes. Third, the proposed model is applicable to recommender systems that can suggest items that match with the user's preference and whose prices coincide with the user's money sense. Most of the existing work on recommender systems focuses on the recommendation of items within the same category, such as movies, books and web pages, where the prices are the same or vary little. However, when recommending diverse items, such as food, electrical products and furniture, price information becomes important since their prices vary widely. Fourth, the inference procedure can be extended further to reduce the computational time by using an online learning approach (Hoffman et al. 2010; Sato et al. 2010) and parallelization (Newman et al. 2007). Finally, we would like to evaluate our model further by applying it to other real purchase data sets.

Appendix A

Derivation of (11)

In this appendix, we give the derivation of (11).

$$\begin{aligned}
&P(z_j = k|\boldsymbol{X}, \boldsymbol{V}, \boldsymbol{Z}_{\setminus j})\\
&\propto P(z_j = k, x_j, v_j|\boldsymbol{X}_{\setminus j}, \boldsymbol{V}_{\setminus j}, \boldsymbol{Z}_{\setminus j})\\
&= P(z_j = k|\boldsymbol{Z}_{\setminus j})P(x_j|\boldsymbol{X}_{\setminus j}, z_j = k, \boldsymbol{Z}_{\setminus j})P(v_j|x_j, \boldsymbol{X}_{\setminus j}, \boldsymbol{V}_{\setminus j}, z_j = k, \boldsymbol{Z}_{\setminus j}),
\end{aligned} \tag{23}$$

The first factor of (23) becomes,

$$\begin{aligned}
&P(z_j = k|\boldsymbol{Z}_{\setminus j})\\
&= \int P(z_j = k|\boldsymbol{\theta}_u)P(\boldsymbol{\theta}_u|\boldsymbol{Z}_{\setminus j})d\boldsymbol{\theta}_u\\
&= \int \theta_{uk}\frac{\Gamma(N_{u\setminus j} + \alpha K)}{\prod_{k'}\Gamma(N_{uk'\setminus j} + \alpha)}\prod_{k'}\theta_{uk'}^{N_{uk'\setminus j}+\alpha-1}d\boldsymbol{\theta}_u\\
&= \frac{\Gamma(N_{u\setminus j} + \alpha K)}{\prod_{k'}\Gamma(N_{uk'\setminus j} + \alpha)}\int \theta_{uk}^{N_{uk'\setminus j}+\alpha}\prod_{k'\neq k}\theta_{uk'}^{N_{uk'\setminus j}+\alpha-1}d\boldsymbol{\theta}_u\\
&= \frac{\Gamma(N_{u\setminus j} + \alpha K)}{\prod_{k'}\Gamma(N_{uk'\setminus j} + \alpha)}\frac{\Gamma(N_{uk\setminus j} + \alpha + 1)\prod_{k'\neq k}\Gamma(N_{uk'\setminus j} + \alpha)}{\Gamma(N_{u\setminus j} + \alpha K + 1)}\\
&= \frac{N_{ku\setminus j} + \alpha}{N_{u\setminus j} + \alpha K},
\end{aligned} \tag{24}$$

where we used $\int\prod_k \theta_k^{\alpha_k-1}d\boldsymbol{\theta} = \frac{\prod_k\Gamma(\alpha_k)}{\Gamma(\sum_k\alpha_k)}$ in the fourth equation, which is the normalizing constant of the Dirichlet distribution, and $\Gamma(x + 1) = x\Gamma(x)$ in the fifth equation. In a similar way, the second factor of (23) becomes,

$$\begin{aligned}
&P(x_j|\boldsymbol{X}_{\setminus j}, z_j = k, \boldsymbol{Z}_{\setminus j})\\
&= \int P(x_j|\boldsymbol{\phi}_k)P(\boldsymbol{\phi}_k|\boldsymbol{X}_{\setminus j}, \boldsymbol{Z}_{\setminus j})d\boldsymbol{\phi}_k\\
&= \int \phi_{kx_j}\frac{\Gamma(N_{k\setminus j} + \beta I)}{\prod_{i'}\Gamma(N_{ki'\setminus j} + \beta)}\prod_{i'}\phi_{ki'}^{N_{ki'\setminus j}+\beta-1}d\boldsymbol{\phi}_k\\
&= \frac{\Gamma(N_{k\setminus j} + \beta I)}{\prod_{i'}\Gamma(N_{ki'\setminus j} + \beta)}\int \theta_{kx_j}^{N_{kx_j\setminus j}+\beta}\prod_{i'\neq x_j}\theta_{ki'}^{N_{ki'\setminus j}+\beta-1}d\boldsymbol{\phi}_k\\
&= \frac{\Gamma(N_{k\setminus j} + \beta I)}{\prod_{i'}\Gamma(N_{ki'\setminus j} + \beta)}\frac{\Gamma(N_{kx_j\setminus j} + \beta + 1)\prod_{i'\neq x_j}\Gamma(N_{ki'\setminus j} + \beta)}{\Gamma(N_{k\setminus j} + \beta I + 1)}\\
&= \frac{N_{kx_j\setminus j} + \beta}{N_{k\setminus j} + \beta I}.
\end{aligned} \tag{25}$$

The third factor of (23) becomes,

$$\begin{aligned}
&P(v_j|x_j, \boldsymbol{X}_{\backslash j}, \boldsymbol{V}_{\backslash j}, z_j = k, \boldsymbol{Z}_{\backslash j})\\
&= \int\int P(v_j|\mu_{kx_j}, \lambda_{kx_j}^{-1}) P(\mu_{kx_j}|\boldsymbol{X}_{\backslash j}, \boldsymbol{V}_{\backslash j}, \lambda_{kx_j}) P(\lambda_{kx_j}|\boldsymbol{X}_{\backslash j}, \boldsymbol{V}_{\backslash j}) d\mu_{kx_j} d\lambda_{kx_j}\\
&= \int\int N(v_j|\mu_{kx_j}, \lambda_{kx_j}^{-1}) N(\mu_{kx_j}|\eta_{kx_j\backslash j}, (\lambda_{kx_j}\gamma_{kx_j})^{-1}) G(\lambda_{kx_j}|a_{kx_j\backslash j}, b_{kx_j\backslash j}) d\mu_{kx_j} d\lambda_{kx_j}\\
&= \frac{\gamma_{kx_j\backslash j}^{\frac{1}{2}} b_{kx_j\backslash j}^{a_{kx_j\backslash j}}}{2\pi\Gamma(a_{kx_j\backslash j})} \int\int \lambda_{kx_j}^{a_{kx_j\backslash j}} \exp\left(-\frac{\lambda_{kx_j}\gamma_{kx_j}}{2}(\mu_{kx_j} - \eta_{kx_j})^2\right) \exp(-\lambda_{kx_j} b_{kx_j}) d\mu_{kx_j} d\lambda_{kx_j}\\
&\propto \frac{\Gamma(a_{kx_j})}{\Gamma(a_{kx_j\backslash j})} \frac{b_{kx_j\backslash j}^{a_{kx_j\backslash j}}}{b_{kx_j}^{a_{kx_j}}} \left(\frac{\gamma_{kx_j\backslash j}}{\gamma_{kx_j}}\right)^{\frac{1}{2}},
\end{aligned} \tag{26}$$

where we used $\int \exp\left(-\frac{\lambda\gamma}{2}(\mu - \eta)^2\right) d\mu = \left(\frac{2\pi}{\lambda\gamma}\right)^{\frac{1}{2}}$ and $\int \lambda^{a-1} \exp(-\lambda b) d\lambda = \frac{\Gamma(a)}{b^a}$ in the third equation, which are the normalizing constant of the Gaussian and Gamma distributions, respectively.

References

Blei DM, Jordan MI (2003) Modeling annotated data. In: SIGIR '03, pp 127–134

Blei DM, Lafferty JD (2006) Dynamic topic models. In: ICML '06, pp 113–120

Blei DM, Ng AY, Jordan MI (2003) Latent Dirichlet allocation. J Mach Learn Res 3:993–1022

Das AS, Datar M, Garg A, Rajaram S (2007) Google news personalization: scalable online collaborative filtering. In: WWW '07, pp 271–280

Fu W, Song L, Xing EP (2009) Dynamic mixed membership block model for evolving networks. In: ICML '09, pp 329–336

Griffiths TL, Steyvers M (2004) Finding scientific topics. Proc Nat Acad Sci 101(1):5228–5235

Hamuro Y, Katoh N, Matsuda Y, Yada K (1998) Mining pharmacy data helps to make profits. Data Min Knowl Disc 2(4):391–398

Hofmann T (1999) Probabilistic latent semantic analysis. In: UAI '99, pp 289–296

Hofmann T (2003) Collaborative filtering via Gaussian probabilistic latent semantic analysis. In: SIGIR '03. ACM Press, New York, pp 259–266

Hoffman M, Blei D, Bach F (2010) Online learning for latent Dirichlet allocation. In: NIPS '10

Iwata T, Saito K, Yamada T (2006) Recommendation methods for extending subscription periods. In: KDD '06, pp 574–579

Iwata T, Watanabe S, Yamada T, Ueda N (2009) Topic tracking model for analyzing consumer purchase behavior. In: IJCAI '09, pp 1427–1432

Iwata T, Yamada T, Sakurai Y, Ueda N (2010) Online multiscale dynamic topic models. In: KDD '10, pp 663–672

Jin X, Zhou Y, Mobasher B (2004) Web usage mining based on probabilistic latent semantic analysis. In: KDD '04, pp 197–205

Li M, Dias BM, Jarman I, El-Deredy W, Lisboa PJ (2009) Grocery shopping recommendations based on basket-sensitive random walk. In: KDD '09, pp 1215–1224

Mimno D, Wallach HM, Naradowsky J, Smith DA, McCallum A (2009) Polylingual topic models. In: EMNLP '09, pp 880–889

Minka T (2000) Estimating a Dirichlet distribution. Technical report. MIT, Cambridge

Minka T, Lafferty J (2002) Expectation-propagation for the generative aspect model. In: UAI '02, pp 352–359

Newman D, Asuncion A, Smyth P, Welling M (2007) Distributed inference for latent Dirichlet allocation. In: NIPS ’07, pp 1081–1088

Nijs VR, Dekimpe MG, Steenkamps JBE, Hanssens DM (2001) The category-demand effects of price promotions. Market Sci 20(1):1–22

Raju JS (1992) The effect of price propotion on variability in product category sales. Market Sci 11(3):207–220

Rosen-Zvi M, Griffiths T, Steyvers M, Smyth P (2004) The author-topic model for authors and documents. In: UAI ’04, pp 487–494

Sato I, Kurihara K, Nakagawa H (2010) Deterministic single-pass algorithm for lda. In: NIPS ’10

Shani G, Heckerman D, Brafman RI (2005) An MDP-based recommender system. J Mach Learn Res 6:1265–1295

Teh YW, Jordan MI, Beal MJ, Blei DM (2006) Hierarchical Dirichlet processes. J Am Stat Assoc 101(476):1566–1581

Teh YW, Newman D, Welling M (2006) A collapsed variational Bayesian inference algorithm for latent Dirichlet allocation. In: NIPS ’06, 1378–1385

Thomas S, Chakravarthy S. (2000) Incremental mining of constrained associations. In: HIPC ’00, 547–558

Wallach HM (2006) Topic modeling: Beyond bag-of-words. In: ICML ’06, 977–984

Wang X, McCallum A (2006) Topics over time: a non-Markov continuous-time model of topical trends. In: KDD ’06, 424–433

Williamson S, Wang C, Heller K, Blei D (2010) The IBP-compound dirichlet process and its application to focused topic modeling. In: ICML ’10, 1151–1158