Topic Tracking Model for Analyzing Consumer Purchase Behavior

Tomoharu Iwata **Shinji Wanatabe** **Takeshi Yamada** **Naonori Ueda**
NTT Communication Science Laboratories,
2-4 Hikaridai, Seika-cho, Soraku-gun, Kyoto, Japan
{iwata,watanabe,yamada,ueda}@cslab.kecl.ntt.co.jp

Abstract

We propose a new topic model for tracking time-varying consumer purchase behavior, in which consumer interests and item trends change over time. The proposed model can adaptively track changes in interests and trends based on current purchase logs and previously estimated interests and trends. The online nature of the proposed method means we do not need to store past data for current inferences and so we can considerably reduce the computational cost and the memory requirement. We use real purchase logs to demonstrate the effectiveness of the proposed method in terms of the prediction accuracy of purchase behavior and the computational cost of the inference.

1 Introduction

Modeling consumer purchase behavior in e-commerce is an important task, because it enables online stores to provide recommendations, personalized advertisements, and perform trend analyses. Recently there has been great interest in analyzing discrete data using topic models, where the discrete data are mainly text corpora but also include human behavior log data [Jin *et al.*, 2004; Das *et al.*, 2007]. A topic model is a hierarchical Bayesian model, in which a user (document) is modeled as a mixture of topics, where a trend in a topic is modeled as a purchase probability distribution over catalog items (words). However, traditional topic models do not fulfill two important requirements for modeling consumer behavior in online stores. The first is the adaptability to changes. Consumer interests and item trends change over time for a variety of reasons. The second is the ease with which incremental model updates can be achieved. Huge amounts of data are accumulated every day in real online stores.

In this paper, we propose a topic model, which we call the *Topic Tracking Model* (TTM), that satisfies the above two requirements. With TTM, interests for each consumer and trends for each topic are sequentially inferred using newly obtained data and previously estimated interests and trends. The computational cost and the memory requirement of TTM are low because past data are not required for the inference and need not be stored.

The persistency of interests differs among consumers. Some consumers significantly change their interests and some do not. Furthermore, it also differs according to periods. There may be periods during which a consumer's interests change greatly and there may also be long periods when these interests remain unchanged. Therefore, we need to infer the interest persistency for each consumer and for each time period. We also infer the trend persistency for each topic and for each time period because it differs among topics and over time according to new releases of items, seasons, and social and economic environments.

Probabilistic Latent Semantic Analysis (PLSA) [Hofmann, 1999] and Latent Dirichlet Allocation (LDA) [Blei *et al.*, 2003] are two representative topic models. However, these models assume that samples are exchangeable, and therefore they cannot capture the dynamics. More recently, topic models with dynamics, such as the Dynamic Topic Model (DTM) [Blei and Lafferty, 2006], Dynamic Mixture Model (DMM) [Wei *et al.*, 2007], and Topic over Time (ToT) [Wang and McCallum, 2006], have been proposed. TTM has an advantage over the other models in its adaptability to changes because TTM models dynamics of both interests and trends, and it infers their persistency for each time interval from the given data.

In the rest of this paper, we assume that the given data are purchase logs. However, our method is applicable to a wide range of human behavior data with timestamps, such as web access log, blog, e-mail and discourse data.

2 Preliminaries

The goal of this paper is to estimate the probability that user u purchases item i at time t, $P(i|u,t)$, which represents the purchase behavior of user u, where $u \in \{1, \cdots, U\}$ is a user index, and $i \in \{1, \cdots, I\}$ is an item index. We assume that t is a discrete variable, and we can set the unit time interval arbitrarily at, for example, one day or one week.

In topic modeling, $P(i|u,t)$ is obtained by the following equation with the assumption of the conditional independence of u and i given latent topic z,

$$P(i|u,t) = \sum_{z=1}^{Z} \phi_{t,u,z}\theta_{t,z,i}, \quad (1)$$

where Z is the number of topics, $\phi_{t,u,z} = P(z|u,t)$ represents the interests of a user, which is the probability that user u is interested in topic z at time t, where $\phi_{t,u,z} \geq 0$, $\sum_z \phi_{t,u,z} = 1$, and $\theta_{t,z,i} = P(i|z,t)$ represents trends in a topic, which is the probability that item i is selected and purchased from topic z at time t, where $\theta_{t,z,i} \geq 0$, $\sum_i \theta_{t,z,i} = 1$.

In LDA, interests $\boldsymbol{\phi}_{t,u} = \{\phi_{t,u,z}\}_{z=1}^{Z}$ are assumed to have a Dirichlet prior as follows:

$$P(\boldsymbol{\phi}_{t,u}|\boldsymbol{\gamma}) \propto \prod_z \phi_{t,u,z}^{\gamma_z - 1}, \tag{2}$$

where $\boldsymbol{\gamma} = \{\gamma_z\}_{z=1}^{Z}$, $\gamma_z > 0$, is a set of Dirichlet parameters. In this assumption, current interests are independent of the past interests, and their dynamics are not incorporated.

3 Topic Tracking Model

To model the dynamics of user interests, we assume that the mean of the user interests at the current time are the same as those at a previous time unless otherwise confirmed by the newly observed data. In particular, we use the following Dirichlet prior, in which the mean of the current interests is the same as the mean of the previous interests $\hat{\boldsymbol{\phi}}_{t-1,u}$, and the precision is $\alpha_{t,u}$,

$$P(\boldsymbol{\phi}_{t,u}|\hat{\boldsymbol{\phi}}_{t-1,u}, \alpha_{t,u}) \propto \prod_z \phi_{t,u,z}^{\alpha_{t,u}\hat{\phi}_{t-1,u,z} - 1}, \tag{3}$$

where Dirichlet parameters in (2) are factorized into the mean and precision $\gamma_{t,u,z} = \alpha_{t,u}\hat{\phi}_{t-1,u,z}$. The precision $\alpha_{t,u}$ represents the interest persistency, which is a measure of how consistently user u maintains her interests at time t compared with those at $t-1$. We estimate $\alpha_{t,u}$ for each time period and for each user because the variability of interests depends on both time and user. Since this is a conjugate prior, the inference is simple as in LDA, which is a topic model without dynamics [Griffiths and Steyvers, 2004].

To model the dynamics of trends, we use the following Dirichlet distribution for the prior of trends $\boldsymbol{\theta}_{t,z} = \{\theta_{t,z,i}\}_{i=1}^{I}$ in the same way as interests,

$$P(\boldsymbol{\theta}_{t,z}|\hat{\boldsymbol{\theta}}_{t-1,z}, \beta_{t,z}) \propto \prod_i \theta_{t,z,i}^{\beta_{t,z}\hat{\theta}_{t-1,z,i} - 1}, \tag{4}$$

where the mean is the previous trends $\hat{\boldsymbol{\theta}}_{t-1,z}$, and the precision is $\beta_{t,z}$. This prior is also conjugate.

Let $\boldsymbol{x}_{t,u} = \{x_{t,u,m}\}_{m=1}^{M_{t,u}}$ be a set of items purchased by user u at time t, where $x_{t,u,m}$ is the mth item purchased by user u at time t, and $M_{t,u}$ is the number of items purchased by user u at time t. TTM assumes the following generative process for $\boldsymbol{X}_t = \{\boldsymbol{x}_{t,u}\}_{u=1}^{U}$, which is a set of all items purchased by U users at time t,

1. For each topic $z = 1, \cdots, Z$:
   (a) Draw trends
   $\boldsymbol{\theta}_{t,z} \sim \mathrm{Dirichlet}(\beta_{t,z}\hat{\boldsymbol{\theta}}_{t-1,z})$
2. For each user $u = 1, \cdots, U$:
   (a) Draw interests
   $\boldsymbol{\phi}_{t,u} \sim \mathrm{Dirichlet}(\alpha_{t,u}\hat{\boldsymbol{\phi}}_{t-1,u})$
   (b) For each purchase $m = 1, \cdots, M_{t,u}$:
      i. Draw topic
      $z_{t,u,m} \sim \mathrm{Multinomial}(\boldsymbol{\phi}_{t,u})$
      ii. Draw item
      $x_{t,u,m} \sim \mathrm{Multinomial}(\boldsymbol{\theta}_{t,z_{t,u,m}})$.

As in LDA, each item $x_{t,u,m}$ is sampled from a topic-specific multinomial distribution.

Figure 1 (a) shows a graphical model representation of TTM, where shaded and unshaded nodes indicate observed and latent variables, respectively.

4 Inference

We estimate the parameters in TTM based on stochastic EM algorithm, in which Gibbs sampling of latent topics and maximum joint likelihood estimation of parameters are alternately iterated [Wallach, 2006].

Let t be the current time. Suppose that we have purchase logs $\boldsymbol{X}_t$ at the current time t, the mean of the previous interests $\hat{\boldsymbol{\Phi}}_{t-1} = \{\hat{\boldsymbol{\phi}}_{t-1,u}\}_{u=1}^{U}$, and the mean of the previous trends $\hat{\boldsymbol{\Theta}}_{t-1} = \{\hat{\boldsymbol{\theta}}_{t-1,z}\}_{z=1}^{Z}$. By using these data and parameters, we estimate current interests $\boldsymbol{\Phi}_t = \{\boldsymbol{\phi}_{t,u}\}_{u=1}^{U}$ and trends $\boldsymbol{\Theta}_t = \{\boldsymbol{\theta}_{t,z}\}_{z=1}^{Z}$ as well as interest persistencies $\boldsymbol{\alpha}_t = \{\alpha_{t,u}\}_{u=1}^{U}$ and trend persistencies $\boldsymbol{\beta}_t = \{\beta_{t,z}\}_{z=1}^{Z}$.

We infer latent topics based on the collapsed Gibbs sampling [Griffiths and Steyvers, 2004], in which the joint distribution of data and latent topics is required. Let $\boldsymbol{Z}_t = \{\boldsymbol{z}_{t,u}\}_{u=1}^{U}$ be a set of latent topics of all users at time t, and $\boldsymbol{z}_{t,u} = \{z_{t,u,m}\}_{m=1}^{M_{t,u}}$. Since we use conjugate priors for parameters $\boldsymbol{\Phi}_t$ and $\boldsymbol{\Theta}_t$, we can integrate out the parameters in the joint distribution as follows:

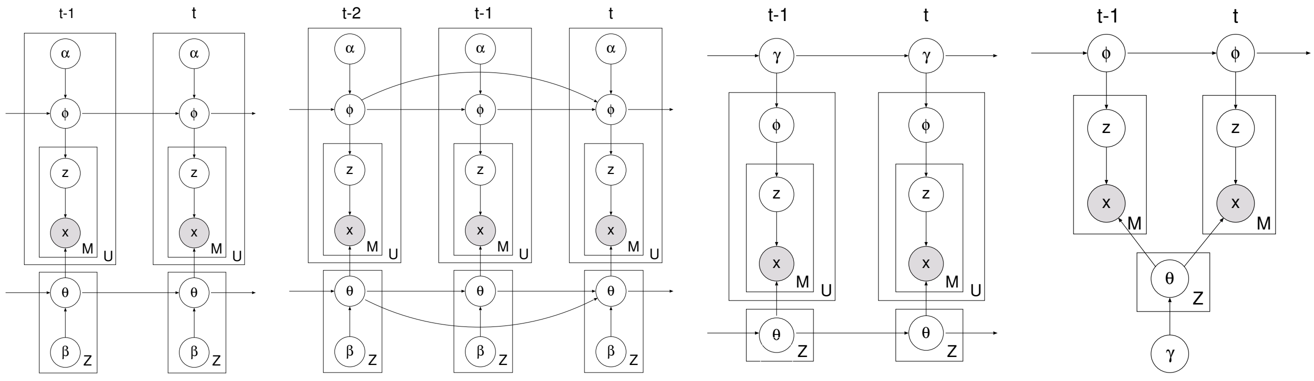
$$\begin{aligned}
P(&\boldsymbol{X}_t, \boldsymbol{Z}_t|\hat{\boldsymbol{\Phi}}_{t-1}, \hat{\boldsymbol{\Theta}}_{t-1}, \boldsymbol{\alpha}_t, \boldsymbol{\beta}_t) \\
&= P(\boldsymbol{Z}_t|\hat{\boldsymbol{\Phi}}_{t-1}, \boldsymbol{\alpha}_t) P(\boldsymbol{X}_t|\boldsymbol{Z}_t, \hat{\boldsymbol{\Theta}}_{t-1}, \boldsymbol{\beta}_t) \\
&= \int P(\boldsymbol{Z}_t, \boldsymbol{\Phi}_t|\hat{\boldsymbol{\Phi}}_{t-1}, \boldsymbol{\alpha}_t) d\boldsymbol{\Phi}_t \\
&\times \int P(\boldsymbol{X}_t, \boldsymbol{\Theta}_t|\boldsymbol{Z}_t, \hat{\boldsymbol{\Theta}}_{t-1}, \boldsymbol{\beta}_t) d\boldsymbol{\Theta}_t \\
&= \prod_u \frac{\Gamma(\alpha_{t,u})}{\prod_z \Gamma(\alpha_{t,u}\hat{\phi}_{t-1,u,z})} \frac{\prod_z \Gamma(n_{t,u,z} + \alpha_{t,u}\hat{\phi}_{t-1,u,z})}{\Gamma(n_{t,u} + \alpha_{t,u})} \\
&\times \prod_z \frac{\Gamma(\beta_{t,z})}{\prod_i \Gamma(\beta_{t,z}\hat{\theta}_{t-1,z,i})} \frac{\prod_i \Gamma(n_{t,z,i} + \beta_{t,z}\hat{\theta}_{t-1,z,i})}{\Gamma(n_{t,z} + \beta_{t,z})},
\end{aligned} \tag{5}$$

where $n_{t,u,z}$ is the number of items purchased by user u at time t that have been assigned to topic z, $n_{t,z,i}$ is the number of times item i has been assigned to topic z at time t, and $\Gamma(x)$ is the gamma function.

The assignment of a latent topic is estimated using the following equation, which is calculated from (5),

$$\begin{aligned}
P(&z_j = k|\boldsymbol{X}_t, \boldsymbol{Z}_{t\backslash j}, \hat{\boldsymbol{\Phi}}_{t-1}, \hat{\boldsymbol{\Theta}}_{t-1}, \boldsymbol{\alpha}_t, \boldsymbol{\beta}_t) \\
&\propto \frac{n_{t,u,k\backslash j} + \alpha_{t,u}\hat{\phi}_{t-1,u,k}}{n_{t,u\backslash j} + \alpha_{t,u}} \frac{n_{t,k,x_j\backslash j} + \beta_{t,k}\hat{\theta}_{t-1,k,x_j}}{n_{t,k\backslash j} + \beta_{t,k}},
\end{aligned} \tag{6}$$

where $j = (t,u,m)$, and $\backslash j$ represents a count that does not include a purchase j.



(a) TTM (b) TTM with long dependences (c) DTM (d) DMM

Figure 1: Graphical model representations of Topic Tracking Model (TTM), Dynamic Topic Model (DTM), and Dynamic Mixture Model (DMM).

The persistency parameters $\boldsymbol{\alpha}_t$ and $\boldsymbol{\beta}_t$ are estimated by maximizing the joint distribution (5). The following updating rules for maximizing the joint distribution are derived by using the bounds in [Minka, 2000],

$$\alpha_{t,u} \leftarrow \alpha_{t,u} \frac{\sum_z \hat{\phi}_{t-1,u,z} A_{t,u,z}}{\Psi(n_{t,u}+\alpha_{t,u}) - \Psi(\alpha_{t,u})} \tag{7}$$

where $A_{t,u,z} = \Psi(n_{t,u,z} + \alpha_{t,u}\hat{\phi}_{t-1,u,z}) - \Psi(\alpha_{t,u}\hat{\phi}_{t-1,u,z})$, and $\Psi(\cdot)$ is the digamma function defined by $\Psi(x) = \frac{\partial \log \Gamma(x)}{\partial x}$, and,

$$\beta_{t,z} \leftarrow \beta_{t,z} \frac{\sum_i \hat{\theta}_{t-1,z,i} B_{t,z,i}}{\Psi(n_{t,z}+\beta_{t,z}) - \Psi(\beta_{t,z})}, \tag{8}$$

where $B_{t,z,i} = \Psi(n_{t,z,i} + \beta_{t,z}\hat{\theta}_{t-1,z,i}) - \Psi(\beta_{t,z}\hat{\theta}_{t-1,z,i})$. By iterating Gibbs sampling with (6), and maximum likelihood estimation with (7) and (8), we can estimate latent topics $\boldsymbol{Z}_t$, and parameters $\boldsymbol{\alpha}_t$ and $\boldsymbol{\beta}_t$.

After the iterations, the means of $\phi_{t,u,z}$ and $\theta_{t,z,i}$ are obtained as follows:

$$\hat{\phi}_{t,u,z} = \frac{n_{t,u,z} + \alpha_{t,u}\hat{\phi}_{t-1,u,z}}{n_{t,u} + \alpha_{t,u}}, \tag{9}$$

$$\hat{\theta}_{t,z,i} = \frac{n_{t,z,i} + \beta_{t,z}\hat{\theta}_{t-1,z,i}}{n_{t,z} + \beta_{t,z}}. \tag{10}$$

These estimates are used to estimate interests and trends at the next time period $t+1$.

TTM uses only current data $\boldsymbol{X}_t$ (not past data) when inferring current interests $\boldsymbol{\Phi}_t$ and trends $\boldsymbol{\Theta}_t$, Therefore, TTM can greatly reduce the computational cost compared with the conventional models that use past data. Moreover, TTM also reduces the memory requirement because past data need not be stored.

5 Capturing long term dependences

In the above discussion, we infer interests $\boldsymbol{\phi}_{t,u}$ using an estimate of the previous interests $\hat{\boldsymbol{\phi}}_{t-1,u}$. However in practice, the interests of today may depend not only on the interests of yesterday but also on the interests of a week ago. Therefore, to capture the long term dependences, we modify the prior of interests $\boldsymbol{\phi}_{t,u}$ so that it can depend on past L interests $\{\hat{\boldsymbol{\phi}}_{t-l,u}\}_{l=1}^{L}$ as follows:

$$P(\boldsymbol{\phi}_{t,u} | \{\hat{\boldsymbol{\phi}}_{t-l,u}, \alpha_{t,u,l}\}_{l=1}^{L}) \propto \prod_z \phi_{t,u,z}^{\sum_l \alpha_{t,u,l}\hat{\phi}_{t-l,u,z}-1}, \tag{11}$$

where the mean is proportional to the weighted sum of the past L interests of the user, and $\alpha_{t,u,l}$ represents how the interests at t are related to the l-previous interests. The information loss is reduced by considering long term dependences. Furthermore, the bias of the inference is reduced by using multiple estimates. When we use only the previous interests, the inference might be biased because it depends on just one estimate.

In a similar fashion, we can also modify the prior of the trends $\boldsymbol{\theta}_{t,u}$ so that it can depend on past L trends $\{\hat{\boldsymbol{\theta}}_{t-l,u}\}_{l=1}^{L}$ in order to make the inference more robust as follows:

$$P(\boldsymbol{\theta}_{t,z} | \{\hat{\boldsymbol{\theta}}_{t-l,z}, \beta_{t,z,l}\}_{l=1}^{L}) \propto \prod_i \theta_{t,z,i}^{\sum_l \beta_{t,z,l}\hat{\theta}_{t-l,z,i}-1}, \tag{12}$$

where the mean is proportional to the weighted sum of the past L trends of the topic, and $\beta_{t,z,l}$ represents how the trends at t are related to the l-previous trends. A graphical model of TTM with long term dependences is shown in Figure 1 (b).

Since the priors in (11) and (12) are still Dirichlet distributions and conjugate priors, the parameters $\boldsymbol{\Phi}_t$ and $\boldsymbol{\Theta}_t$ in the joint distribution can be integrated out as in (5). Therefore, the inference can be achieved by using a stochastic EM algorithm in the same way with the model described in the previous section. Each latent topic is sampled according to the following probability,

$$P(z_j = k | \boldsymbol{X}_t, \boldsymbol{Z}_{t\backslash j}, \{\hat{\boldsymbol{\Phi}}_{t-l}, \hat{\boldsymbol{\Theta}}_{t-l}, \boldsymbol{\alpha}_{t,l}, \boldsymbol{\beta}_{t,l}\}_{l=1}^{L})$$
$$\propto \frac{n_{t,u,k\backslash j} + \sum_l \alpha_{t,u,l}\hat{\phi}_{t-l,u,k}}{n_{t,u\backslash j} + \sum_l \alpha_{t,u,l}} \frac{n_{t,k,x_j\backslash j} + \sum_l \beta_{t,k,l}\hat{\theta}_{t-l,k,x_j}}{n_{t,k\backslash j} + \sum_l \beta_{t,k,l}}, \tag{13}$$

where $\boldsymbol{\alpha}_{t,l} = \{\alpha_{t,u,l}\}_{u=1}^{U}$, and $\boldsymbol{\beta}_{t,l} = \{\beta_{t,z,l}\}_{z=1}^{Z}$. The update rule of weight $\alpha_{t,u,l}$ is as follows:

$$\alpha_{t,u,l} \leftarrow \alpha_{t,u,l} \frac{\sum_z \hat{\phi}_{t-l,u,z} A'_{t,u,z}}{\Psi(n_{t,u} + \sum_{l'} \alpha_{t,u,l'}) - \Psi(\sum_{l'} \alpha_{t,u,l'})}, \quad (14)$$

where $A'_{t,u,z} = \Psi(n_{t,u,z} + \sum_l \alpha_{t,u,l}\hat{\phi}_{t-l,u,z}) - \Psi(\sum_l \alpha_{t,u,l}\hat{\phi}_{t-l,u,z})$. The update rule of weight $\beta_{t,z,l}$ is as follows:

$$\beta_{t,z,l} \leftarrow \beta_{t,z,l} \frac{\sum_i \hat{\theta}_{t-l,z,i} B'_{t,z,i}}{\Psi(n_{t,z} + \sum_{l'} \beta_{t,z,l'}) - \Psi(\sum_{l'} \beta_{t,z,l'})}, \quad (15)$$

where $B'_{t,z,i} = \Psi(n_{t,z,i} + \sum_l \beta_{t,z,l}\hat{\theta}_{t-l,z,i}) - \Psi(\sum_l \beta_{t,z,l}\hat{\theta}_{t-l,z,i})$. The means of $\phi_{t,u,z}$ and $\theta_{t,z,i}$ are obtained as follows:

$$\hat{\phi}_{t,u,z} = \frac{n_{t,u,z} + \sum_l \alpha_{t,u,l}\hat{\phi}_{t-l,u,z}}{n_{t,u} + \sum_l \alpha_{t,u,l}}, \quad (16)$$

$$\hat{\theta}_{t,z,i} = \frac{n_{t,z,i} + \sum_l \beta_{t,z,l}\hat{\theta}_{t-l,z,i}}{n_{t,z} + \sum_l \beta_{t,z,l}}. \quad (17)$$

These estimates are efficiently used for the inference from $t+1$ to $t+L$.

6 Related work

Topic models with dynamics have been proposed recently. These include the Dynamic Topic Model (DTM) [Blei and Lafferty, 2006], Dynamic Mixture Model (DMM) [Wei *et al.*, 2007], and Topic over Time (ToT) [Wang and McCallum, 2006]. A graphical model of DTM is shown in Figure 1 (c). DTM analyzes the time evolution of topics in document collections, in which a document is assumed to have only one timestamp. Therefore, DTM does not consider the dynamics in each document (user), and it cannot be used for modeling user behavior dynamics, which is the goal of this paper. Since DTM uses a Gaussian distribution for the dynamics, the inference is intractable owing to the nonconjugacy of the Gaussian and multinomial distributions, and it requires complicated variational approximations. In contrast to DTM, the inference of TTM is relatively simple because it uses Dirichlet distributions that are conjugate for multinomial distributions. A graphical model of DMM is shown in Figure 1 (d). DMM only considers a single dynamic sequence of documents, which corresponds to a single user over time. On the other hand, TTM considers multiple users with different dynamics as well as the dynamics of trends for multiple items in each topic, which is essential for analyzing consumer purchase behavior. Furthermore, DMM assumes that the interest persistency is fixed and known. Therefore, it is not appropriate for purchase logs, in which distributions sometimes change greatly, and are sometimes stable. ToT needs all the samples over time for the inference. Therefore, it cannot be updated sequentially, and is not appropriate for data that are continuously accumulated. Singular Value Decomposition (SVD) is also used for analyzing multiple timeseries [Papadimitriou *et al.*, 2005] as well as topic models. However, since SVD implicitly assumes Gaussian noise, it is not appropriate for discrete data [Hofmann, 1999], such as text corpora and purchase logs.

TTM is related to the Kalman filter [Welch and Bishop, 1995] even though TTM deals with discrete variables and the Kalman filter deals with continuous variables. In the Kalman filter, since both the transition and emission distributions are Gaussian, and the conjugate distribution of the Gaussian is also Gaussian, the inference is tractable. On the other hand, with topic models, since the emission distribution is the multinomial, and the conjugate distribution of the multinomial is the Dirichlet, the inference in the same way is infeasible. DTM is inferred as with the Kalman filter by mapping the Gaussian to the multinomial using the softmax function. In contrast, TTM is inferred by assuming that the transition distribution is the Dirichlet, in which the mean is the previous mean of the multinomial, and the variance is adaptively estimated from the given data.

7 Experiments

We evaluated TTM using two real purchase log data sets consisting of movie and cartoon data. The movie data are the logs of a movie downloading service from May 14th to August 31st, 2007, in which the numbers of users, items, and transactions are 70,122, 7,469, and 11,243,935, respectively. The cartoon data are the logs of a cartoon downloading service for cell phones from January 1st to May 31st, 2006, in which the numbers of users, items, and transactions are 143,212, 206, and 12,642,505, respectively. We considered cartoons with different volumes to be the same item. We omitted items and users that appeared fewer than ten times. We set the unit time interval at one day.

For the evaluation measurement, we employed the N-best prediction accuracy of purchase items. We used data until the previous day $t-1$ as training data to infer the model, and predicted purchase items at day t as test data. The probability that user u purchases item i at time t was calculated using the estimated parameters at the previous day, $\hat{\boldsymbol{\Phi}}_{t-1}$ and $\hat{\boldsymbol{\Theta}}_{t-1}$, as follows:

$$P(x=i|u,t;\hat{\boldsymbol{\Phi}}_{t-1},\hat{\boldsymbol{\Theta}}_{t-1}) = \sum_z \hat{\phi}_{t-1,u,z}\hat{\theta}_{t-1,z,i}. \quad (18)$$

The N-best accuracy represents the percentage that the purchased items are contained in the set of N-highest $P(x{=}i|u,t;\hat{\boldsymbol{\Phi}}_{t-1},\hat{\boldsymbol{\Theta}}_{t-1})$ items.

In order to demonstrate the effectiveness of modeling the dynamics of interests and trends, we compared TTM with three topic models, LDAall, LDAonline, and LDAone, which are based on LDA. LDAall is an LDA that uses all the past data for inference. LDAonline is an online learning extension of LDA, in which the parameters are estimated using those of the previous day and the newly obtained data [Banerjee and Basu, 2007]. LDAone is an LDA that uses just the previous day data for inference. For a fair comparison, the hyper parameters in these LDAs were optimized using stochastic EM as described in [Wallach, 2006].

The N-best accuracies for movie and cartoon data sets averaged over time are shown in Table 1. TTM1 and TTM10

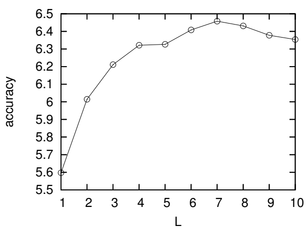
Table 1: Average N-best accuracies (%) over time. The digit in the bracket is the standard deviation.

(a) movie

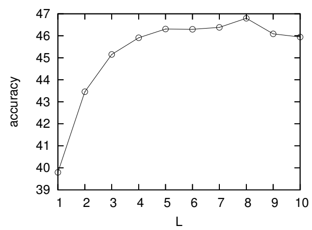
N	LDAall	LDAonline	LDAone	TTM1	TTM10
1	1.21 (0.61)	1.08 (0.54)	1.91 (0.78)	2.22 (0.91)	**2.46** (0.92)
2	2.18 (0.79)	2.00 (0.78)	3.52 (1.22)	3.99 (1.33)	**4.47** (1.36)
3	3.06 (1.04)	2.81 (1.02)	5.04 (1.64)	5.60 (1.75)	**6.35** (1.85)
4	3.90 (1.27)	3.56 (1.24)	6.24 (1.90)	6.82 (2.01)	**7.82** (2.15)
5	4.70 (1.51)	4.26 (1.44)	7.37 (2.20)	7.92 (2.26)	**9.20** (2.42)

(b) cartoon

N	LDAall	LDAonline	LDAone	TTM1	TTM10
1	27.0 (3.3)	26.0 (3.5)	24.8 (4.5)	26.8 (4.2)	**30.5** (3.4)
2	37.3 (3.6)	35.1 (4.2)	32.4 (4.9)	34.2 (4.5)	**39.9** (3.5)
3	43.7 (3.9)	41.1 (4.8)	37.2 (5.3)	39.8 (4.6)	**45.9** (3.3)
4	48.5 (4.0)	45.8 (5.1)	40.9 (5.3)	44.5 (4.6)	**50.6** (3.2)
5	52.4 (4.2)	49.6 (5.4)	44.1 (5.4)	48.5 (4.6)	**54.4** (3.0)



(a) movie

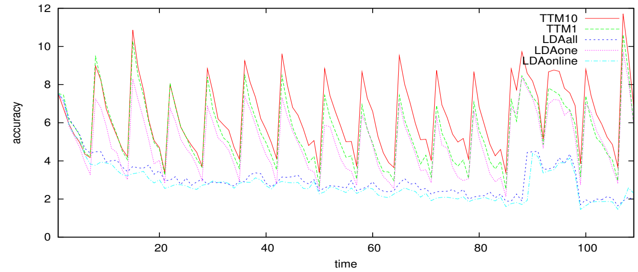


(b) cartoon

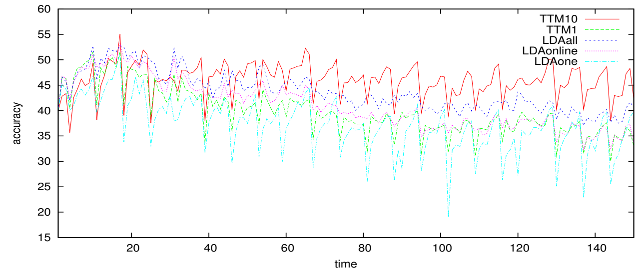
Figure 2: Average three-best accuracies (%) of TTM with different dependent lengths L.

represent the proposed topic tracking models, which use parameters of one- and ten-previous days, or $L = 1$ and $L = 10$, respectively. For both of the data sets, the highest accuracies are achieved by TTM10, and this result indicates that TTM10 can appropriately predict consumer behavior by considering dynamics and efficiently using information in the past data. The accuracies achieved by LDAall and LDAonline are low because they do not consider the dynamics. The reason for the low accuracy of LDAone is that it uses only current data and ignores the past information. The accuracies achieved by LDAall and LDAonline for the cartoon data are higher than that of LDAone because the cartoon data has smaller dynamics than the movie data. The high accuracies of TTM10 in both of the data sets that have different dynamics represent its high adaptability.

Figure 2 shows the average three-best accuracies of TTM with different dependent lengths L. The accuracy increased as L increased. This result implies that TTM became robust by using long term dependences. The accuracy saturated at around $L = 7$, and then slightly decreased possibly due to overfitting. Therefore, we can achieve high accuracy with these data sets by storing estimated parameters for a week. The three-best accuracies for movie and cartoon data sets for each day are shown in Figure 3. The accuracies varied with a period of seven days because new items are released every week, and the accuracies when new items were released were



(a) movie



(b) cartoon

Figure 3: Three-best accuracies (%) for each day.

Table 2: Average computational time (sec).

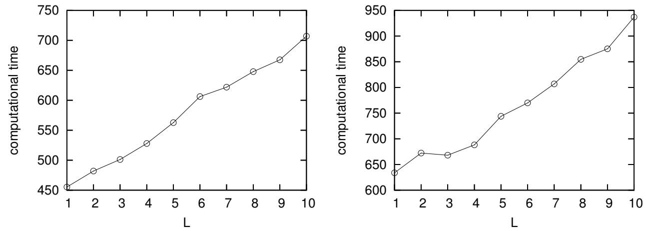
	LDAall	LDAonline	LDAone	TTM1	TTM10
movie	12,380.3	523.7	503.2	455.3	707.0
cartoon	12,528.4	663.2	674.6	633.9	936.9

low.

Table 2 shows the daily average computational time when using a computer with a Xeon5355 2.66GHz CPU. The computational time for LDAonline, LDAone and TTM are roughly the same since they use the same number of samples for inference. The computational time for LDAall is long although its accuracy was the second highest for the cartoon data set since LDAall uses all samples for inference. Moreover, LDAall requires more memory than TTM. TTM10 processed data that consisted of over 100,000 transactions per day in about 12 minutes, and it can deal with the huge amounts of data that are observed in real online stores. Figure 4 shows the average computational time for TTM with different dependent lengths L. The computational time increases linearly with L. Note that the time for TTM is much smaller than that for LDAall even TTM with $L = 10$.

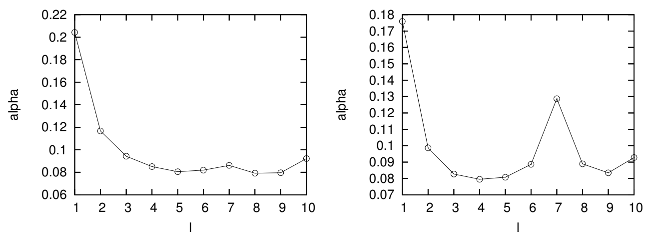
Figures 5 and 6, respectively, show the average α and β with different l values of TTM10. The sum of the values for each day and for each user or topic are normalized to one. The parameters decrease as l increases in the movie data set. This result implies that recent interests and trends are more informative as regards estimating current interests and trends, which is an intuitive result. In the cartoon data set, the parameters at $l = 7$ are high because the same cartoons with different volumes are released weekly.

Figure 7 shows some examples of estimated parameter β of TTM1 for each day with the movie data. The trends in a topic (red) have a high dependence on the previous trends,



(a) movie (b) cartoon

Figure 4: Average computational time (sec) of TTM with different dependent lengths L.



(a) movie (b) cartoon

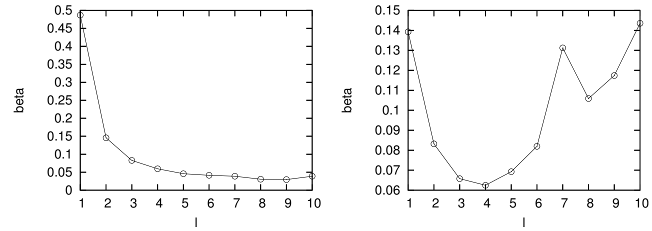
Figure 5: Average normalized α with different lengths l of TTM10.

and the trends in a topic (green) have a low dependence on the previous trends. Even in a topic, the dependence changes day by day. For example, in a topic (red), the dependence on the day of a new release is low and those of the other days are high. In this way, TTM can adapt flexibly to changes.

8 Conclusion

In this paper, we proposed a probabilistic consumer purchase behavior model, called the Topic Tracking Model (TTM), for tracking the interests of individual consumers and trends in each topic. We confirmed experimentally that TTM could estimate purchase behavior appropriately and process large amounts of data.

In future work, we could determine the unit time interval and the length of dependence automatically from the given data. The number of topics would be automatically inferred by extending TTM to a nonparametric Bayesian model such as the Dirichlet process mixture model [Teh *et al.*, 2006]. We will evaluate the proposed method further by comparing it with other methods for modeling dynamics, such as hidden Markov models, dynamic topic models, and dynamic mixture models.



(a) movie (b) cartoon

Figure 6: Average normalized β with different lengths l of TTM10.

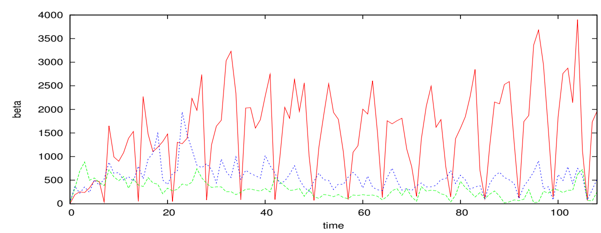


Figure 7: Examples of β of TTM1 for each day with the movie data.

References

[Banerjee and Basu, 2007] Arindam Banerjee and Sugato Basu. Topic models over text streams: A study of batch and online unsupervised learning. In *SDM '07*, 2007.

[Blei and Lafferty, 2006] David M. Blei and John D. Lafferty. Dynamic topic models. In *ICML '06*, pages 113–120, 2006.

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

[Das *et al.*, 2007] Abhinandan S. Das, Mayur Datar, Ashutosh Garg, and Shyam Rajaram. Google news personalization: scalable online collaborative filtering. In *WWW '07*, pages 271–280, 2007.

[Griffiths and Steyvers, 2004] Thomas L. Griffiths and Mark Steyvers. Finding scientific topics. *Proceedings of the National Academy of Sciences*, 101 Suppl 1:5228–5235, 2004.

[Hofmann, 1999] Thomas Hofmann. Probabilistic latent semantic analysis. In *UAI '99*, pages 289–296, 1999.

[Jin *et al.*, 2004] Xin Jin, Yanzan Zhou, and Bamshad Mobasher. Web usage mining based on probabilistic latent semantic analysis. In *KDD '04*, pages 197–205, 2004.

[Minka, 2000] Thomas Minka. Estimating a Dirichlet distribution. Technical report, MIT, 2000.

[Papadimitriou *et al.*, 2005] Spiros Papadimitriou, Jimeng Sun, and Christos Faloutsos. Streaming pattern discovery in multiple time-series. In *VLDB '05*, pages 697–708, 2005.

[Teh *et al.*, 2006] Y. W. Teh, M. I. Jordan, M. J. Beal, and D. M. Blei. Hierarchical Dirichlet processes. *Journal of the American Statistical Association*, 101(476):1566–1581, 2006.

[Wallach, 2006] Hanna M. Wallach. Topic modeling: Beyond bag-of-words. In *ICML '06*, pages 977–984, 2006.

[Wang and McCallum, 2006] Xuerui Wang and Andrew McCallum. Topics over time: a non-Markov continuous-time model of topical trends. In *KDD '06*, pages 424–433, 2006.

[Wei *et al.*, 2007] Xing Wei, Jimeng Sun, and Xuerui Wang. Dynamic mixture models for multiple time-series. In *IJCAI '07*, pages 2909–2914, 2007.

[Welch and Bishop, 1995] Greg Welch and Gary Bishop. An introduction to the Kalman filter. Technical report, University of North Carolina at Chapel Hill, 1995.