Toward Personalized Context Recognition for Mobile Users: A Semisupervised Bayesian HMM Approach

BAOXING HUAI, ENHONG CHEN, and HENGSHU ZHU, University of Science and Technology of China
HUI XIONG, Rutgers University
TENGFEI BAO and QI LIU, University of Science and Technology of China
JILEI TIAN, Nokia

The problem of mobile context recognition targets the identification of semantic meaning of context in a mobile environment. This plays an important role in understanding mobile user behaviors and thus provides the opportunity for the development of better intelligent context-aware services. A key step of context recognition is to model the personalized contextual information of mobile users. Although many studies have been devoted to mobile context modeling, limited efforts have been made on the exploitation of the sequential and dependency characteristics of mobile contextual information. Also, the latent semantics behind mobile context are often ambiguous and poorly understood. Indeed, a promising direction is to incorporate some domain knowledge of common contexts, such as "waiting for a bus" or "having dinner," by modeling both labeled and unlabeled context data from mobile users because there are often few labeled contexts available in practice. To this end, in this article, we propose a sequence-based semisupervised approach to modeling personalized context for mobile users. Specifically, we first exploit the Bayesian Hidden Markov Model (B-HMM) for modeling context in the form of probabilistic distributions and transitions of raw context data. Also, we propose a sequential model by extending B-HMM with the prior knowledge of contextual features to model context more accurately. Then, to efficiently learn the parameters and initial values of the proposed models, we develop a novel approach for parameter estimation by integrating the Dirichlet Process Mixture (DPM) model and the Mixture Unigram (MU) model. Furthermore, by incorporating both user-labeled and unlabeled data, we propose a semisupervised learning-based algorithm to identify and model the latent semantics of context. Finally, experimental results on real-world data clearly validate both the efficiency and effectiveness of the proposed approaches for recognizing personalized context of mobile users.

Categories and Subject Descriptors: H.2.8.d [**Information Technology and Systems**]: Database Applications—*Data mining*; I.2.6 [**Artificial Intelligence**]: Learning

General Terms: Algorithms, Experimentation

Additional Key Words and Phrases: Context recognition, hidden Markov model

---

This work was supported in part by grants from the National Science Foundation for Distinguished Young Scholars of China (Grant No. 61325010), the Natural Science Foundation of China (NSFC, Grant No. 71329201), the International Science and Technology Cooperation Plan of Anhui Province (Grant No. 1303063008), and the Anhui Provincial Natural Science Foundation (Grant No. 1408085QF110). This work was also partially supported by grants from the National Science Foundation (NSF) via grant numbers CCF-1018151 and IIS-1256016.
Author's addresses: B. Huai, E. Chen, H. Zhu, T. Bao, and Q. Liu, School of Computer Science and Technology, University of Science and Technology of China, Hefei, Anhui 230026, China; emails: bxhuai@mail.ustc.edu.cn; cheneh@ustc.edu.cn; zhs@mail.ustc.edu.cn; tfbao92@mail.ustc.edu.cn; qiliuql@ustc.edu.cn; H. Xiong, Management Science and Information Systems Department Rutgers, the State University of New Jersey, Newark, NJ07102, USA; email: hxiong@rutgers.edu; J. Tian, Nokia Research Center, Beijing 100010, China; email: jilei.tian@nokia.com.

DOI: http://dx.doi.org/10.1145/2629504

ACM Reference Format:
Baoxing Huai, Enhong Chen, Hengshu Zhu, Hui Xiong, Tengfei Bao, Qi Liu, and Jilei Tian. 2014. Toward personalized context recognition for mobile users: A semisupervised Bayesian HMM approach. ACM Trans. Knowl. Discov. Data 9, 2, Article 10 (September 2014), 29 pages.
DOI: http://dx.doi.org/10.1145/2629504

1. INTRODUCTION

Advances in the sensing capabilities of smart mobile devices have enabled the accumulation of rich contextual and interaction information from mobile users through device logs. A distinct property of user interactions with mobile devices is that they are usually associated with volatile contexts that always contain rich semantic meanings, such as waiting for a bus, driving a car, or doing shopping. Therefore, recognizing the semantic meanings of such contextual information plays an important role in understanding user habits and thus opening a venue for the development of intelligent, context-aware services such as context-aware recommendations and habit-based user segmentation [Liampotis et al. 2012; Anagnostopoulos et al. 2007; Lemlouma and Layaïda 2004; Ma et al. 2012]. Indeed, a key step for context recognition is modeling the context of the raw data collected from mobile users. This is also a fundamental research problem in leveraging the rich contextual information of mobile users.

In the literature, although there are some previous studies on mobile context modeling [Bao et al. 2012; Abowd et al. 1997; Schilit et al. 1994], few of them pay attention to the sequential and dependency characteristics of mobile contextual information. However, we notice that mobile contexts are often mutually dependent, and the semantic meaning of a context may be related to other adjacent contexts. For example, if we only observe a context of Tom (e.g., {(*Day Period: Evening*), (*Location: On the way to home*)}), it is hard to uncover its latent semantics. However, if we also observe another adjacent context (e.g., {(*Day Period: Evening*), (*Location: Work place*)}), it can be inferred that *Tom is going home after work*, which is the latent semantics behind these contexts. Therefore, it is very useful (and also a crucial challenge in determining how) to take the sequential and dependency characteristics into consideration during contextual modeling. Moreover, the latent semantic meanings behind mobile contexts are often ambiguous and poorly understood. Thus, it is appealing to incorporate some domain knowledge of common contexts, such as *waiting for a bus* or *having dinner* into the context modeling process. Actually, some of the semantic information can be obtained through user interactions with mobile devices, such as the usage of life-logging software and diary applications [Belimpasakis et al. 2009; Teraoka 2011; Rawassizadeh et al. 2012]. These user-labeled contexts present an intuitive way to ease the process of mobile context modeling. However, in practice, there are often few labeled but numerous unlabeled context data. Therefore, how to model contextual information through both labeled and unlabeled context data collected from mobile users is another crucial challenge to deal with.

To fill this crucial void, in this article, we propose a sequence-based semisupervised approach to modeling personalized context for mobile users. Specifically, we first exploit the Bayesian Hidden Markov Model (B-HMM) [Goldwater and Griffiths 2007] to model context in the form of probabilistic distributions and transitions of raw context data. Also, we propose a novel approach by extending B-HMM with the prior knowledge of contextual features, namely HMMC, for modeling context more effectively. Then, to efficiently learn the parameters and initial values of the proposed models, we also develop a novel method for parameter estimation by integrating the Dirichlet Process Mixture (DPM) model and the Mixture Unigram (MU) model. Moreover, to incorporate both labeled and unlabeled contexts collected from mobile users into the process of context modeling, we develop a semisupervised learning-based algorithm to identify and model the latent and ambiguous semantic meanings of context, which are referred as ***Context Topics*** in this article. In particular, we developed a software *ActivityLogger* for collecting rich contextual information and semantic labels from mobile users. Finally, experimental results on the collected real-world datasets clearly validate the effectiveness and efficiency of the proposed approaches for recognizing the personalized contexts of mobile users. We summarize our contributions in this article as follows:

- First, we propose two novel graphical models, B-HMM and HMMC, for context modeling. Both of them can accurately model the personalized context of mobile users. To the best of our knowledge, this article is the first attempt to model contextual information of mobile users by taking the sequential and dependency characteristics of contextual information into consideration.
- Second, we propose a novel approach to effectively estimate the number of latent context topics of contexts by integrating the DPM model and the MU model. This approach can also be used to accelerate the training process of our contextual models.
- Third, we develop a straightforward but efficient semisupervised learning-based algorithm to identify and model the latent semantics of context. This algorithm integrates both labeled and unlabeled contextual data from mobile users into the process of context modeling.
- Finally, we develop a tool, *ActivityLogger,* to collect rich contextual information and context topic labels from mobile users. Moreover, we perform extensive experiments based on the collected real-world data for validating both the effectiveness and efficiency of our contextual models.

Overview. The rest of this article is organized as follows. In Section 2, we introduce related works. Section 3 shows the preliminaries of mobile context modeling. In Section 4, we present our context modeling approaches by exploiting B-HMM. Section 5 presents the semisupervised learning-based algorithm. In Section 6, we show the experimental results. Finally, Section 7 concludes the work.

2. RELATED WORK

In general, the related works of this study can be grouped into four categories. The first category includes research studies about **context recognition**. For example, Himberg et al. [2001] studied the problem of context recognition in mobile scenarios based on unsupervised segmentation of time series. They proposed two greedy dynamic algorithms to find optimal *k*-segmentation for a given cost function. There are also some works on context recognition based on audio and video data processing [Eronen et al. 2006; Chu 2008]. Specifically, Eronen et al. [2006] considered context recognition as a task of automatic context classification by using only acoustic information, where the context is just a location (e.g., a restaurant or a marketplace). In Korpipää et al. [2003], authors applied Naive Bayesian networks to recognize the contexts of mobile users with respect to their daily activities, where the context data are mainly from audio sensors. However, most of these works only focused on specific contextual information, such as places, activities, and the like. In fact, abundant contextual information can be collected by smart mobile devices. Therefore, it motivates us to propose a more comprehensive and scalable approach for modeling the context of mobile users.

In the second category, researchers focus on **modeling mobile context**. For example, some researchers tried to model context from GPS data [Liao et al. 2007; Zheng et al. 2008], accelerometer data [Ravi et al. 2005; Nham et al. 2008], or multiple dimension context data [Bao et al. 2012; Cao et al. 2010]. Specifically, Liao et al. [2007] attempted to infer an individual's transportation routine given the user's raw GPS data by leveraging a dynamic Bayesian network. Zheng et al. [2008] exploited several supervised learning approaches for modeling users' raw GPS data. Eagle and Pentland [2009] proposed using the eigenvector of user behaviors for modeling individual user's context and inferring community affiliations within the subject's social network. Bao et al. [2012] extended two topic models (i.e., LDA and MU models) to model personalized contexts for mobile users. Indeed, most of the existing works use either supervised or unsupervised methods. Each has its benefits and drawbacks: Although the supervised learning approach can provide more flexibility, it depends on labeled training data that cannot be obtained easily from the real lives of mobile users. By contrast, an unsupervised approach can model contexts more flexibly because it does not need labeled data. However, unsupervised approaches cannot explicitly reveal the semantic meanings of contexts. Meanwhile, some of the context semantics can be labeled by users using life-logging software or the diary applications found in smart devices. Intuitively, these user-labeled contexts provide an intuitive way to ease the process of mobile context modeling. Thus, in this article, we propose a semisupervised learning framework to model contexts for mobile users.

In the third category, we introduce some previous works on **mining the context logs of mobile users**. For example, some researchers have proposed leveraging context logs for mobile app classification [Zhu et al. 2012a] and mining mobile users' activities [Peng et al. 2012]. Yu et al. [2012] proposed a novel, personalized, context-aware recommender system by analyzing mobile user's context logs. The proposed approach is based on the Latent Dirichlet Allocation topic model and is scalable for multiple contextual features. Furthermore, Zhu et al. [2012b] proposed a uniform framework for personalized context-aware recommendation that can integrate both context independency and dependency assumptions. The framework can mine user's personal context-aware preferences for mobile app recommendations from the context logs of many mobile users. Li et al. [2012] proposed a learning- based approach for inferring the status of high-energy-consuming sensors according to the outputs of software-based sensors and the physical sensors that are necessary to work continuously to support the basic functions of mobile devices. Moreover, some researchers also proposed mining behavior patterns of mobile users from context logs [Cao et al. 2010; Ma et al. 2012], which can be used to support context-aware services. However, most existing works on mining context logs (including mobile context modeling) do not consider the sequential and dependency characteristics of the context records. Therefore, we propose a novel sequential approach based on a Bayesian Hidden Markov model for modeling the context of mobile users.

Finally, the last category concerns **Hidden Markov Models** (HMMs), which have been successfully applied to problems in a variety of fields, such as signal processing and speech recognition [Rabiner 1989; Juang and Rabiner 1991], biometrics [Fredkin and Rice 1992; Leroux and Puterman 1992], genetics [Churchill 1989; Liu et al. 1999], economics [Hamilton 1989; Albert and Chib 1993], and search log mining [Cao et al. 2009]. For much of their history, HMMs have been implemented by using recursive algorithms developed for parameter estimation [Baum et al. 1970], which are viewed as "black boxes" by many statisticians. In recent years, some researchers proposed using Bayesian methods to simulate HMM parameters from posterior distribution, which can provide a more scalable and stable process of parameter estimation for HMM. For example, a novel B-HMM model has been proposed for part-of-speech tagging [Goldwater and Griffiths 2007], and Guha et al. [2008] adopt B-HMM to model array Comparative Genomic Hybridization (CGH) data. pievatolo et al. [2012] attempted to detect imperfect debugging from the possible introduction of bugs during software debugging through a B-HMM-based method. Compared with the traditional Maximum-Likelihood

Table I. A Toy Context Log Collected by Mobile Devices

Timestamp	Context record	User Label
t_1	{(Day name: Monday), (Is holiday: No), (Time range: AM6:00-7:00)), (Profile: Silent), (Battery Level: Full), (Cell ID: $id_1, \ldots$}	Sleeping
t_2	{(Day name: Monday), (Is holiday: No), (Time range: AM6:00-7:00)), (Profile: Silent), (Battery Level: Full), (Cell ID: $id_1, \ldots$}	Sleeping
t_3	{(Day name: Monday), (Is holiday: No), (Time range: AM6:00-7:00)), (Profile: Silent), (Battery Level: Full), (Cell ID: $id_1, \ldots$}	Sleeping
......	......	
t_{229}	{(Day name: Monday), (Is holiday: No), (Time range: AM10:00-11:00), (Profile: General), (Battery Level: High), (Cell ID: $id_2, \ldots$}	Null
t_{230}	{(Day name: Monday), (Is holiday: No), (Time range: AM10:00-11:00), (Profile: General), (Battery Level: High), (Cell ID: $id_2, \ldots$}	Null
......	......	
t_{343}	{(Day name: Monday), (Is holiday: No), (Time range: AM11:00-12:00), (Profile: Meeting), (Battery Level: High), (Cell ID: $id_2, \ldots$}	Meeting
t_{344}	{(Day name: Monday), (Is holiday: No), (Time range: AM11:00-12:00), (Profile: Meeting), (Battery Level: High), (Cell ID: $id_2, \ldots$}	Meeting

Estimation (MLE)-based HMM learning solution, B-HMM can directly maximize the probability of hidden variables given the observed data by integrating over all possible parameter values rather than searching for an optimal set of parameter values. Therefore, previous studies have clearly proved its robustness and scalability in modeling sequential data that contain domain priori [Leggetter and Woodland 1995]. To this end, we propose to leverage B-HMM for modeling context data.

3. PRELIMINARIES OF MODELING MOBILE CONTEXT

The context-collecting software of smart devices can capture the rich context data of mobile users through multiple context sensors, and these data are recorded in context logs. For example, Table I shows a toy context log that contains several ***context records***, and each context record consists of a timestamp and the detailed contextual information recorded at that time. A context consists of several ***contextual features*** (e.g., Day name, Time range) and their corresponding values (e.g., Saturday, AM8:00-9:00), which can be annotated as ***contextual feature-value pairs***. Indeed, the context collection software predefines a set of contextual features whose values should be collected; however, a context record may miss the values of some contextual features for mainly two reasons. First, these values are not always available (e.g., GPS signal cannot be received when the device is indoors or the corresponding sensor has been closed). Second, users have no related operations (e.g., there are no *Event* values when users do not use their mobile devices). Thus, only the contextual feature-value pairs whose values are not missing are recorded in context logs.

We can discover some latent semantic meanings (i.e., ***context topics***) from the context records in the context logs of mobile users. For example, suppose Table I shows a part of Bob's context log, where the real cell IDs are replaced by id_1 and id_2 for privacy concerns. From this table, we can see that during time *AM6:00-AM7:00*, the mobile phone is in charging and the profile type is *Silent*, which might imply the semantic meaning *"Bob was sleeping."* Similarly, during the time *AM11:00-AM12:00*, the profile was *Meeting*; thus, considering that the cell ID represents a business place, the context might indicate that *"Bob was having a meeting."* Moreover, we observe that the latent context topics of a context may depend on previous contexts. For example, in Table I, the context topic of the context record at t_2 (i.e., $\boldsymbol{r}_{t_2}$) may depend on the context record at

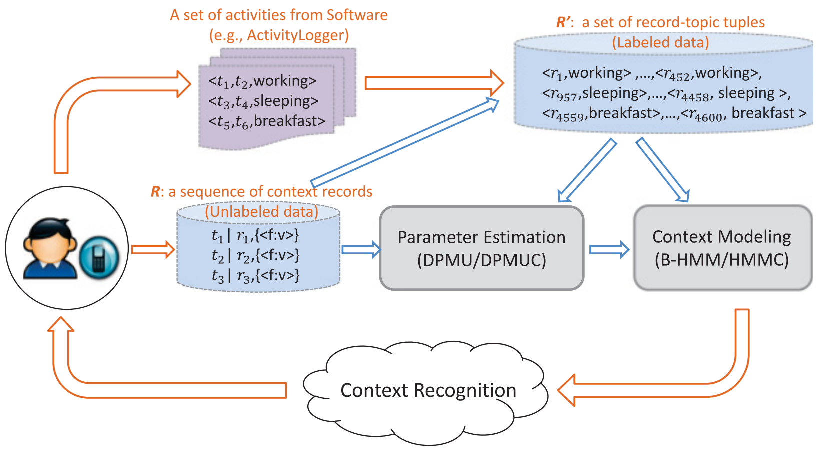


Fig. 1. The framework overview of our novel context modeling approaches.

t_1 (i.e., $\boldsymbol{r}_{t_1}$). Thus, if we only observe the context record $\boldsymbol{r}_{t_2}$, it is hard to decide whether Bob was *"Sleeping"* or *"Working"* since both activities can happen in such a context. However, if we also observe and know that the context topic of the context record $\boldsymbol{r}_{t_1}$ is *"Sleeping,"* it is more likely that Bob is *"Sleeping"* in the context of $\boldsymbol{r}_{t_2}$.

Therefore, if we can uncover these latent topics of contexts, we can provide some context-aware services for mobile users. For example, if Bob is in a context that represents *"Go shopping,"* an intelligent application may recommend to him some popular stores and discounts. Indeed, users can also manually label their contexts when using some life-logging software or diary applications [Belimpasakis et al. 2009; Teraoka 2011; Rawassizadeh et al. 2012], and these labels can be recorded in their context logs. Intuitively, these semantic labels of contexts can describe user habits quite accurately and thus are very helpful for modeling the context of mobile users. However, users will not manually label all their contexts in practice; thus many of the context records may not contain semantic labels (e.g., "Null" in Table I). To solve this problem, in we propose a semisupervised learning algorithm to model contexts for mobile users by incorporating both labeled and unlabeled data.

Figure 1 shows the framework overview of our novel context modeling approaches. Specifically, we first collect many labeled and unlabeled context data through the context collection software. Then we propose two sequential models (i.e., B-HMM and HMMC) for modeling the context of mobile users. Particularly, to guarantee better modeling performance, we propose a novel parameter estimation approach based on the DPM and MU models. Finally, we integrate all the labeled and unlabeled data into a semisupervised algorithm for improving the performance of context recognition.

4. MODELING PERSONALIZED CONTEXT FOR MOBILE USERS

In this section, we first present how to utilize the B-HMM [Goldwater and Griffiths 2007] to model mobile context, and then we propose a novel context modeling approach (i.e., HMMC model) by extending B-HMM with the prior knowledge of contextual features. Before that, we first introduce several basic notations used in this article.

The context records of a mobile user can be represented by a set of sequential records $R = \{\boldsymbol{r}_1, \boldsymbol{r}_2, \ldots, \boldsymbol{r}_t, \ldots\}$, where $\boldsymbol{r}_t$ denotes the context record with timestamp t. Indeed, one context record $\boldsymbol{r}_t \in R$ can be represented by a set of contextual feature-value pairs $\boldsymbol{r}_t = \{p_{t,1}, p_{t,2}, \ldots\}$, where each $p_{t,j}$ is the j-th contextual feature-value pair in context record $\boldsymbol{r}_t$. Specifically, the contextual feature-value pair $p_{t,j} = (f : v)$ (e.g., *Battery*

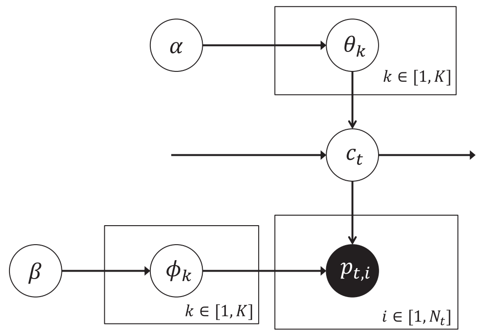


Fig. 2. The graphical representation of the B-HMM for modeling context.

Level: High), where f is the corresponding contextual feature (i.e., *Battery Level*), v is the value of f (i.e., *High*). Moreover, in this article, we assume that each user may have K latent context topics (e.g., *"Working"*) and use c_t to denote the topic of the context record $\boldsymbol{r_t}$. Therefore, the objective of context modeling is to identify the latent context topics from the context logs of mobile users.

4.1. Bayesian Hidden Markov Model for Context Modeling

B-HMM is a variation of the classic HMM [Rabiner 1989], which was originally proposed for part-of-speech tagging. Previous studies have clearly proved its robustness and scalability in modeling sequential data with various domain knowledge [Goldwater and Griffiths 2007; Guha et al. 2008]. To this end, we propose to leverage B-HMM for modeling context data by considering the sequential and dependency characteristics of mobile contextual information.

Specifically, B-HMM has the structure of a standard bi-gram hidden Markov model that contains symmetric Dirichlet priors over the transition and emission distributions for modeling the sequential context records. Therefore, the dependency relationships of B-HMM in our problem can be represented as follows:

$$\begin{aligned} c_t | c_{t-1}, \Theta &\sim Mult(\theta_{c_{t-1}}) \\ p_{t,j} | c_t, \Phi &\sim Mult(\phi_{c_t}) \\ \theta_{c_{t-1}} | \alpha &\sim Dirichlet(\alpha) \\ \phi_{c_t} | \beta &\sim Dirichlet(\beta) \end{aligned}$$

where $c_t|c_{t-1}, \Theta \sim Mult(\theta_{c_{t-1}})$ means c_t follows $Mult(\theta_{c_{t-1}})$ based on given c_{t-1} and Θ [Teh et al. 2006; Blei and Lafferty 2006]. c_t indicates the topic of the context record with timestamp t (i.e., $\boldsymbol{r_t}$), and $p_{t,j}$ is the j-th contextual feature-value pair of the context record $\boldsymbol{r_t}$. $\theta_{c_{t-1}}$ is the topic transition distribution of context record $\boldsymbol{r_t}$ when the topic of the previous context record is c_{t-1}, and ϕ_{c_t} is the output emission distribution of the contextual feature-value pairs of the context c_t. Particularly, both Θ and Φ follow the Dirichlet distribution with parameters α and β. Based on these settings, B-HMM can be treated as a generative model, and Figure 2 shows its graphical representation.

According to Figure 2, the B-HMM assumes that a contextual feature-value pair $p_{t,i}$ of a context record $\boldsymbol{r_t}$ is generated as follows. First, a prior transition distribution of context topics (i.e., θ) is generated from a prior Dirichlet distribution α. Second, a prior output distribution of contextual feature-value pairs (i.e., ϕ) is generated from a prior Dirichlet distribution β. Third, a context topic c_t is generated from $\theta_{c_{t-1}}$ with respect to the previous context topic c_{t-1}. Finally, a contextual feature-value pair $p_{t,i}$ is generated from the distribution ϕ_{c_t}. Note that, according to the definition of Dirichlet distribution, both α and β can be represented by parameter vectors $\alpha \in \mathbb{R}_{+}^{K}$ and $\beta \in \mathbb{R}_{+}^{J}$, respectively, where K is the number of latent context topics and J is the number of unique contextual feature-value pairs.

Given the hyperparameters α and β in this generative model, we can calculate the joint distribution of all observations and hidden variables in B-HMM by the following equation:

$$P(\boldsymbol{r_t}, c_t, \Theta, \Phi | c_{t-1}, \alpha, \beta) = P(\Theta|\alpha)P(\Phi|\beta)P(c_t|c_{t-1}, \Theta)\left(\prod_{i=1}^{N_t} P(p_{t,i}|c_t, \Phi)\right), \tag{1}$$

where N_t is the number of contextual feature-value pairs in context record $\boldsymbol{r_t}$, and

$$P(c_t|c_{t-1}, \Theta) = P(c_t|c_{t-1}, \theta_{c_{t-1}}), \tag{2}$$

$$P(p_{t,i}|c_t, \Phi) = P(p_{t,i}|c_t, \phi_{c_t}). \tag{3}$$

Therefore, the likelihood of a set of context records R can be calculated as follows:

$$L(R) = \int \prod_{k=1}^{K} P(\theta_k|\alpha) \prod_{t=1}^{|R|} P(c_t|c_{t-1}, \theta_{c_{t-1}}) d\Theta \int \prod_{k=1}^{K} P(\phi_k|\beta) \prod_{t=1}^{|R|} \prod_{i=1}^{N_t} P(p_{t,i}|c_t, \phi_{c_t}) d\Phi. \tag{4}$$

The process of training B-HMM is to learn the proper latent variables Θ and Φ to maximize the likelihood in Equation (4). However, the likelihood representation is so complex that it may be not feasible to calculate the corresponding parameters directly. Alternatively, according to some previous studies [Bao et al. 2012], here, we propose to use a popular iterative approach, namely Gibbs sampling [Heinrich 2005; Resnik and Hardisty 2010], for approximately estimating the parameters. Specifically, Gibbs sampling is a widely applicable Markov chain Monte Carlo algorithm that can be seen as a special case of the Metropolis-Hastings algorithm. In our problem, the algorithm begins with a random assignment of context topics to context records with predefined parameters α and β for initializing the state of the Markov chain. In each iteration of the chain, the method will re-estimate the conditional probability by assigning a context topic to each context record, which is conditioned on the assignment of all other records except the current one. Then, according to those conditional probabilities, a new assignment of records to context topics will be scored as a new state of the Markov chain. Finally, after enough rounds of iteration, the assignment will converge, which means that every context record is assigned stable context topics.

From the joint distribution in Equation (1), we can derive the full conditional distribution for each context record $\boldsymbol{r_t}$ (i.e., the updating equation from which the Gibbs sampler draws the context topic to c_t). As shown in Appendix A.1, we have

$$P(c_t|\boldsymbol{c}_{\neg \boldsymbol{r_t}}, R) \propto P(\boldsymbol{r_t}|c_t, \boldsymbol{c}_{\neg \boldsymbol{r_t}}, \Lambda_{\neg \boldsymbol{r_t}})P(c_t|\boldsymbol{c}_{\neg \boldsymbol{r_t}})P(c_{t+1}|\boldsymbol{c}_{\neg(\boldsymbol{r_t}, \boldsymbol{r_{t+1}})}), \tag{5}$$

where $\neg \boldsymbol{r_t}$ means removing context record $\boldsymbol{r_t}$ from R, $\boldsymbol{c}_{\neg \boldsymbol{r_t}}$ denotes the context topics of all context records except $\boldsymbol{r_t}$, Λ is the set of all contextual feature-value pairs in R, and $\Lambda_{\neg \boldsymbol{r_t}}$ denotes all contextual feature-value pairs in R except those in $\boldsymbol{r_t}$.

Furthermore, according to the Gibbs sampling rules, we can estimate each multiplier in Equation (5) by

$$P(\boldsymbol{r_t}|c_t, \boldsymbol{c}_{\neg \boldsymbol{r_t}}, \Lambda_{\neg \boldsymbol{r_t}}) = \frac{\prod_{j=1}^{J} \prod_{i=1}^{N_{t,j}} (n_{\neg \boldsymbol{r_t}, k, j} + \beta_j + i)}{\prod_{i=1}^{N_t} \left(\sum_{j=1}^{J} n_{\neg \boldsymbol{r_t}, k, j} + \beta_j + i \right)}, \tag{6}$$

$$P(c_t = k|\boldsymbol{c}_{\neg \boldsymbol{r_t}}) = \frac{n_{\neg \boldsymbol{r_t}, (c_{t-1}, k)} + \alpha_k}{n_{\neg \boldsymbol{r_t}, (c_{t-1}, *)} + \sum_{k'=1}^{K} \alpha_{k'}}, \tag{7}$$

$$P(c_{t+1}|\boldsymbol{c}_{\neg(\boldsymbol{r_t},\boldsymbol{r_{t+1}})}) = \frac{n_{\neg(\boldsymbol{r_t},\boldsymbol{r_{t+1}}),(k,c_{t+1})} + \mathbf{I}(c_{t-1} = k = c_{t+1}) + \alpha_k}{n_{\neg(\boldsymbol{r_t},\boldsymbol{r_{t+1}}),k} + \mathbf{I}(c_{t-1} = k) + \sum_{k'=1}^{K} \alpha_{k'}}, \tag{8}$$

where $N_{t,j}$ is the number of times the j-th context feature-value pair of Λ appears in r_t, $n_{\neg\boldsymbol{r_t},k,j}$ is the number of times the j-th context feature-value pair of Λ appears in the context record that is assigned to context topic k across all context records except for $\boldsymbol{r_t}$, $n_{\neg\boldsymbol{r_t},(c_{t-1},k)}$ is the number of times context records are assigned to context topic k when the former record (i.e., $\boldsymbol{r}_{t-1}$) is assigned to context c_{t-1} except for $\boldsymbol{r_t}$, and $n_{\neg\boldsymbol{r_t},(c_{t-1},*)} = \sum_{k'=1}^{K} n_{\neg\boldsymbol{r_t},(c_{t-1},k')}$. $n_{\neg(\boldsymbol{r_t},\boldsymbol{r_{t+1}}),k}$ indicates the number of times context records are assigned to context topic k, except $\boldsymbol{r_t}$ and $\boldsymbol{r_{t+1}}$. Moreover, $\mathbf{I}(\mathbf{x})$ is an indicator function whose value equals to 1 when x is true and 0 otherwise. In particular, we present the derivation in Appendix A.3 and A.4. After enough rounds of iteration, the assignment will converge; thus, we can estimate the parameters Θ and Φ by

$$\theta_{c_{t-1},k} = \frac{n_{c_{t-1},k} + \alpha_k}{\sum_{k'} n_{c_{t-1},k'} + \alpha_{k'}}, \quad \phi_{c_t,p_{t,j}} = \frac{n_{c_t,p_{t,j}} + \beta_{p_{t,j}}}{\sum_{p'} n_{c_t,p'} + \beta_{p'}}, \tag{9}$$

where $\theta_{c_{t-1},k} = P(c_t|c_{t-1}, \theta_{c_{t-1}})$ is the probability that current context record $\boldsymbol{r_t}$ is assigned to topic k (i.e, $c_t = k$) when the topic of the previous context record is c_{t-1}, and $\phi_{c_t,p_{t,j}} = P(p_{t,j}|c_t, \phi_{c_t})$ is the probability that contextual feature-value pair $p_{t,j}$ will appear in the current context record with topic c_t.

4.1.1. Context Topic Inference by B-HMM. After learning the B-HMM for context modeling, another task is inferring the context topic c_t for a given new context record $\boldsymbol{r_t}$. Specifically, the problem is to calculate the probability $P(c_t = k|c_{t-1}, \boldsymbol{r_t}, \Theta, \Phi)$ for each context topic $k \in 1 : K$ and find the context topic k^* that satisfies

$$k^* = \arg\max_k P(c_t = k|c_{t-1}, \boldsymbol{r_t}, \Theta, \Phi). \tag{10}$$

To solve this problem, we propose the context topic inference as follows:

$$\begin{aligned} P(c_t = k|c_{t-1}, \boldsymbol{r_t}, \Theta, \Phi) &\propto P(c_t, \boldsymbol{r_t}|c_{t-1}, \Theta, \Phi) \\ &\propto P(c_t|c_{t-1}, \Theta) P(\boldsymbol{r_t}|c_t, \Phi) \\ &\propto P(c_t|c_{t-1}, \Theta) \times \prod_{i=1}^{N_t} P(p_{t,i}|c_t, \Phi), \end{aligned} \tag{11}$$

where $P(c_t|c_{t-1}, \Theta) = \theta_{c_{t-1},k}$, and $P(p_{t,i}|c_t, \Phi) = \phi_{c_t,p_i}$. Particularly, for the context record $\boldsymbol{r_t}$ that does not have previous record $\boldsymbol{r}_{t-1}$, we set $\theta_{c_{t-1},k}$ to $\frac{1}{K}$.

4.2. Bayesian Hidden Markov Model with Prior Knowledge of Contextual Features

Although B-HMM can model the context data in an intuitive way, from some real-world observations we find the generation of contextual feature-value pairs is determined not only by latent context topics, but also by their internal contextual features. For example, contextual information can only be obtained when users open corresponding sensors, and GPS information often cannot be obtained due to the lack of signal (e.g., when users are in underground subways). To solve such problems, we propose a novel approach to model context data by extending the B-HMM with prior knowledge of contextual features, namely HMMC.

The graphical representation of HMMC is shown in Figure 3, and it assumes that the generation of context records not only depends on prior context topic distribution, but also on contextual feature distribution. According to the graphical model of the HMMC, a context record $\boldsymbol{r_t}$ is generated as follows. First, a prior transition distribution

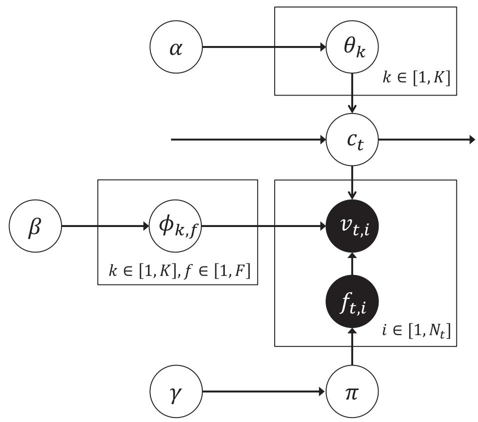


Fig. 3. The graphical representation of the HMMC for modeling context.

of context topics (i.e., θ) is generated from a prior Dirichlet distribution α. Second, a global prior distribution of contextual features (i.e., π) is generated from a prior Dirichlet distribution γ. Third, a prior output distribution of contextual feature-value pairs (i.e., ϕ) is generated from a Dirichlet distribution β. Fourth, a context topic c_t is generated from $\theta_{c_{t-1}}$ with respect to the previous context topic c_{t-1}. Finally, a contextual feature $f_{t,i}$ is generated from π, and the value of $f_{t,i}$, denoted as $v_{t,i}$, is generated from the distribution $\phi_{c_t, f_{t,i}}$.

Given parameters α, β, γ and the context topic of the previous record c_{t-1} in the HMMC, we can calculate the joint probability $P(\boldsymbol{r_t}, c_t, \pi, \Theta, \Phi | c_{t-1}, \alpha, \beta, \gamma)$ by

$$\begin{aligned} P(\boldsymbol{r_t}, c_t, \pi, \Theta, \Phi | c_{t-1}, \alpha, \beta, \gamma) \;=\;& P(\Theta|\alpha) P(\Phi|\beta) P(\pi|\gamma) \\ & \times P(c_t | c_{t-1}, \Theta) \left(\prod_{i=1}^{N_t} P(v_{t,i} | c_t, f_{t,i}, \Phi) P(f_{t,i} | \pi) \right), \end{aligned} \tag{12}$$

where we have

$$P(v_{t,i} | c_t, f_{t,i}, \Phi) = P(v_{t,i} | c_t, f_{t,i}, \phi_{c_t, f_{t,i}}) = P(v_{t,i} | \phi_{c_t, f_{t,i}}). \tag{13}$$

Therefore, the likelihood of all context records R can be represented by

$$\begin{aligned} L(R) = & \int \prod_{j=1}^{F} P(\pi_j|\gamma) d\pi \times \int \prod_{k=1}^{K} P(\theta_k|\alpha) \prod_{t=1}^{|R|} P(c_t | c_{t-1}, \theta_{c_{t-1}}) d\Theta \\ & \times \int \prod_{k=1}^{K} \prod_{j=1}^{F} P(\phi_{k,j}|\beta) \prod_{t=1}^{|R|} \prod_{i=1}^{N_t} P(v_{t,i} | c_t, f_{t,i}, \phi_{c_t, f_{t,i}}) P(f_{t,i}|\pi) d\Phi, \end{aligned} \tag{14}$$

where F indicates the number of unique contextual features. The process of training HMMC is to learn the proper latent variables Θ, π, and Φ to maximize the likelihood in Equation (14). Similar to B-HMM, we also propose to utilize Gibbs sampling to estimate the parameters for HMMC. Specifically, as shown in Appendix A.2, the corresponding

Gibbs sampler of c_t is given as follows:

$$P(c_t|\boldsymbol{c}_{\neg \boldsymbol{r_t}}, R) \propto P(\boldsymbol{v_t}|c_t, \boldsymbol{c}_{\neg \boldsymbol{r_t}}, F, V_{\neg \boldsymbol{r_t}})P(c_t|\boldsymbol{c}_{\neg \boldsymbol{r_t}})P(c_{t+1}|\boldsymbol{c}_{\neg(\boldsymbol{r_t},\boldsymbol{r_{t+1}})}), \tag{15}$$

where $P(c_t|\boldsymbol{c}_{\neg \boldsymbol{r_t}})$ and $P(c_{t+1}|\boldsymbol{c}_{\neg(\boldsymbol{r_t},\boldsymbol{r_{t+1}})})$ can be estimated similarly to Equation (7) and Equation (8). $P(\boldsymbol{v_t}|c_t, \boldsymbol{c}_{\neg \boldsymbol{r_t}}, F, V_{\neg \boldsymbol{r_t}})$ can be computed by

$$P(\boldsymbol{v_t}|c_t = k, \boldsymbol{c}_{\neg \boldsymbol{r_t}}, F, V_{\neg \boldsymbol{r_t}}) = \frac{\prod_{v\in \boldsymbol{v_t}} \prod_{j=1}^{N_{t,v}} (n_{\neg \boldsymbol{r_t},k,f,v} + \beta_v + j)}{\prod_{i=1}^{N_t} \sum_{v'\in V_{f_i}} (n_{\neg \boldsymbol{r_t},k,f_i,v'} + \beta_{v'} + i)}, \tag{16}$$

where V_{f_i} is the set of contextual values of feature f_i, $N_{t,v}$ is the number of times value v occurs in $\boldsymbol{r_t}$, N_t is the number of contextual feature-value pairs in $\boldsymbol{r_t}$, and $n_{\neg \boldsymbol{r_t},k,f,v}$ is the number of times the contextual feature-value pair $(f : v)$ is assigned to context topic k. After sufficient rounds of iteration, the assignment will converge; thus, the parameter Θ can be estimated in the same way as in Equation (9). Particularly, the parameters Φ and π can be calculated by:

$$\phi_{c_t,f,v} = \frac{n_{c_t,f,v} + \beta_{f,v}}{\sum_{v'\in V_f} n_{c_t,f,v'} + \beta_{f,v'}}, \quad \pi_f = \frac{n_f + \gamma_f}{\sum_{f'}(n_{f'} + \gamma_{f'})}, \tag{17}$$

where n_f is the number of times that contextual feature f appears in all context records R.

4.2.1. Context Topic Inference by HMMC. After learning the HMMC for context modeling, another task is to infer the context topic c_t for a given new context record $\boldsymbol{r_t}$. Specifically, the problem is to calculate the probability $P(c_t = k|c_{t-1}, \boldsymbol{r_t}, \Theta, \Phi, \pi)$ for each context topic $k \in 1 : K$ and find the topic semantic k^* that satisfies

$$k^* = \arg\max_k P(c_t = k|c_{t-1}, \boldsymbol{r_t}, \Theta, \Phi, \pi). \tag{18}$$

To solve this problem, we propose the context topic inference as follows:

$$\begin{aligned} P(c_t = k|c_{t-1}, \boldsymbol{r_t}, \Theta, \Phi, \pi) &\propto P(c_t, \boldsymbol{r_t}|c_{t-1}, \Theta, \Phi, \pi) \\ &\propto P(c_t|c_{t-1}, \Theta)P(\boldsymbol{r_t}|c_t, \Phi, \pi) \\ &\propto P(c_t|c_{t-1}, \Theta) \times \prod_{i=1}^{N_t} P(v_{t,i}|f_{t,i}, c_t, \Phi)P(f_{t,i}|\pi), \end{aligned} \tag{19}$$

where $P(c_t|c_{t-1}, \Theta) = \theta_{c_{t-1},k}$, $P(v|f, c_t, \Phi) = \phi_{c_t,f,v}$, and $P(f|\pi) = \pi_f$. Similar to B-HMM, in HMMC, for the context record $\boldsymbol{r_t}$ that does not have previous record $\boldsymbol{r}_{t-1}$, we have $\theta_{c_{t-1},k} = \frac{1}{K}$.

4.3. Estimating the Number of Context Topics

Both B-HMM and HMMC need a predefined parameter K to determine the number of context topics. A commonly used approach for estimating K in context modeling is leveraging perplexity [Azzopardi et al. 2003; Blei et al. 2003]. In this approach, we should first predefine a number range $\mathbb{K} = [K_{min}, K_{max}]$ and then calculate the perplexity of the dataset for each $K \in \mathbb{K}$ until we obtain the best K with the lowest perplexity. However, it is not a trivial work to decide an accurate parameter range $\mathbb{K}$; thus, this perplexity based approach usually needs to try different K many times before getting the best one, which is very inefficient in practice. To solve this problem, in this article, we propose to extend the DPM model for estimating the best number of context topics by integrating the MU model [Nigam et al. 2000] and the Mixture Unigram on Context (MUC) model [Bao et al. 2012]. DPM is a mixture model with an infinite number of mixture components [Teh et al. 2006]; it has been well studied in previous literatures for determining the number of clusters [Antoniak 1974; Ishwaran and James 2001; Neal 2000; Huang et al. 2012]. Specifically, given a randomly selected initial number of clusters, DPM can automatically learn the best number of clusters. Thus, DPM is an efficient method to be used for parameter estimating. Indeed, we can understand DPM through a metaphor called the Chinese Restaurant Process (CRP) [Teh et al. 2006].

The hierarchical Bayesian specification of the DPM in our problem is given as follows:

$$
\begin{aligned}
G|\alpha, G_0 &\sim DP(\alpha, G_0) \\
\theta_{c_t}|G &\sim G, t = 1, 2, \ldots\ldots, N \\
c_t|\theta_{c_t} &\sim F(c_t|\theta_{c_t}), t = 1, 2, \ldots, N,
\end{aligned}
$$

where each latent context topic of context record c_t is drawn from one of the K multinomial distributions. Let θ_{c_t} be the parameter of distribution from which the latent topic c_t is generated. Because the number of context topics is unknown, to guarantee that the number of context topics can grow with data, we assume that c_t follows a general mixture model in which θ_{c_t} is generated from a distribution G. The $DP(\alpha, G_0)$ is a Dirichlet Process (**DP**) with a base distribution G_0 and a positive scaling parameter. Intuitively, α can be seen as the inverse variance, and G_0 is the mean of the DP.

The **MU** model is a single-topic-based topic model that can be used to model mobile contexts. Specifically, MU assumes that a context record $\boldsymbol{r_t}$ is generated as follows. Given K context topics and Λ contextual feature-value pairs, to generate the feature-value pair $p_{t,i}$ in $\boldsymbol{r_t}$, a context topic c_t is first generated from a prior topic distribution for the context records R. Then, pair $p_{t,i}$ is generated from the prior record distribution for c_t. Its extension (i.e., the **MUC** model) assumes that a context record is generated by a prior contextual feature distribution and a prior context topic distribution together. Specifically, given K context topics and F contextual features, a context record $\boldsymbol{r_t}$ is generated as follows. A global prior context topic distribution θ is generated from a prior Dirichlet distribution α first, and then a prior contextual feature distribution π is generated from a prior Dirichlet distribution γ. Then, a context topic c_t is generated from θ. Last, a contextual feature $f_{t,i}$ is generated from π, and the value of $f_{t,i}$ (i.e., $v_{t,i}$), is generated from the distribution $\phi_{c_t, f_{t,i}}$.

Moreover, both the prior context topic distribution and the prior context record distribution in the MU and MUC models follow the Dirichlet distribution. In addition, context topic distributions (i.e., θ) in MU and MUC are not similar to those of B-HMM and HMMC proposed in our article, whereas θ in B-HMM/HMMC is the context topic transition distribution. Indeed, MU and B-HMM have a similar parameter Φ, whereas MUC and HMMC have a similar parameter Φ.

Therefore, the key issue is calculating the probability of choosing an existing context topic and the probability of generating a new context topic for the current context record. To calculate these probabilities and decide the best number of context topics, we propose two novel Gibbs sampling-based approaches. Specifically, one is DPMU, arrived at by combining both DPM and MU model for B-HMM, and the other is DPMUC, arrived at by combining both DPM and MUC models for HMMC.

Specifically, the approach starts with a random assignment of K_0 context topics to context records, where K_0 is the initial number of context topics. In each of the following iterations, we first count the number of unique context topics in all context records (i.e., K) and then calculate the two probabilities according to the Gibbs rules to assign context topics for each context record. Particularly, for the B-HMM model, we have a DPMU model as follows:

$$\begin{aligned}P(c_t|\boldsymbol{c}_{\neg \boldsymbol{r_t}}, R) &\propto P(c_t, \boldsymbol{c}_{\neg \boldsymbol{r_t}}, R) = P(c_t, \boldsymbol{c}_{\neg \boldsymbol{r_t}}, V)\\ &= P(\Lambda|c_t, \boldsymbol{c}_{\neg \boldsymbol{r_t}})P(c_t|\boldsymbol{c}_{\neg \boldsymbol{r_t}})\\ &\propto P(\boldsymbol{r_t}|c_t, \boldsymbol{c}_{\neg \boldsymbol{r_t}}, \Lambda_{\neg \boldsymbol{r_t}})P(c_t|\boldsymbol{c}_{\neg \boldsymbol{r_t}});\end{aligned} \tag{20}$$

then, the conditional distributions of $\boldsymbol{r_t}$ and c_t are as follows:

$$P(\boldsymbol{r_t}|c_t, \boldsymbol{c}_{\neg \boldsymbol{r_t}}, V_{\neg \boldsymbol{r_t}}) = \begin{cases} \frac{\prod_{j=1}^{J}\prod_{i=1}^{N_{t,j}}(n_{\neg \boldsymbol{r_t},k,j}+\beta_j+i)}{\prod_{i=1}^{N_t}(\sum_{j=1}^{J} n_{\neg \boldsymbol{r_t},k,j}+\beta_j+i)}, & if\ c_t \le K \\ \frac{\prod_{j=1}^{J}\prod_{i=1}^{N_{t,j}}(\beta_j+i)}{\prod_{i=1}^{N_t}\sum_{j=1}^{J}(\beta_j+i)}, & if\ c_t = K+1 \end{cases}, \tag{21}$$

$$P(c_t|\boldsymbol{c}_{\neg \boldsymbol{r_t}}) = \begin{cases} k, k \le K, & with\ probability\ \frac{n_{\neg \boldsymbol{r_t},k}}{N-1+\alpha} \\ new\ context\ topic\ K+1, & with\ probability\ \frac{\alpha}{N-1+\alpha} \end{cases}. \tag{22}$$

For the HMMC model, the Gibbs sampler of the DPMUC model is given as follows:

$$\begin{aligned}P(c_t|\boldsymbol{c}_{\neg \boldsymbol{r_t}}, R) &\propto P(c_t, \boldsymbol{c}_{\neg \boldsymbol{r_t}}, R)\\ &= P(c_t, \boldsymbol{c}_{\neg \boldsymbol{r_t}}, V, F)\\ &= P(V|c_t, \boldsymbol{c}_{\neg \boldsymbol{r_t}}, F)P(c_r|\boldsymbol{c}_{\neg \boldsymbol{r_t}})P(F)\\ &\propto P(\boldsymbol{v_t}|c_t, \boldsymbol{c}_{\neg \boldsymbol{r_t}}, F, V_{\neg \boldsymbol{r_t}})P(c_r|\boldsymbol{c}_{\neg \boldsymbol{r_t}}),\end{aligned} \tag{23}$$

where the calculation of $P(c_t|\boldsymbol{c}_{\neg \boldsymbol{r_t}})$ is the same as Equation (22), and $P(\boldsymbol{v_t}|c_t, \boldsymbol{c}_{\neg \boldsymbol{r_t}}, F, V_{\neg \boldsymbol{r_t}})$ can be computed by

$$P(\boldsymbol{v_t}|c_t, \boldsymbol{c}_{\neg \boldsymbol{r_t}}, F, V_{\neg \boldsymbol{r_t}}) = \begin{cases} \frac{\prod_{v\in \boldsymbol{v_t}}\prod_{j=1}^{N_{t,v}}(n_{\neg \boldsymbol{r_t},k,f,v}+\beta_v+j)}{\prod_{i=1}^{N_t}\sum_{v\in V_{f_i}}(n_{\neg \boldsymbol{r_t},k,f_i,v}+\beta_v+i)}, & if\, c_t \le K \\ \frac{\prod_{v\in \boldsymbol{v_t}}\prod_{j=1}^{N_{t,v}}(\beta_v+j)}{\prod_{i=1}^{N_t}\sum_{v\in \Lambda_{f_i}}(\beta_v+i)}, & if\, c_t = K+1 \end{cases}. \tag{24}$$

Finally, after several rounds of Gibbs sampling, the number of context topics will be stable; thus, we can get the best parameter K^* for context modeling. In practice, we can first randomly choose the value of K_0 to initiate the Gibbs sampling and finish the sampling if the difference of model likelihoods between two iterations is less than 1%.

Last, but not least, after obtaining the best number of context topics by DPMU/DPMUC, we can reuse the sampling results in the last round of implementing DPMU/DPMUC to initiate the training process of B-HMM/HMMC. Experimental results show that reusing the results of DPMU/DPMUC can not only dramatically accelerate the training process of B-HMM/HMMC, but also can improve the performance (e.g., likelihood) of context modeling in B-HMM/HMMC.

5. SEMISUPERVISED LEARNING FOR CONTEXT MODELING

Indeed, learning in B-HMM, HMMC, and DPMU/DPMUC models can be treated as an unsupervised process. However, the latent context topics behind mobile contexts are often ambiguous and are poorly understood during context recognition [Bao et al. 2012]. Fortunately, in practice, some context records can be manually labeled by mobile users through their interactions, such as when using life-logging software and diary applications. Intuitively, if we can incorporate such semantic labels into the process of context modeling, the learning and recognition results will be dramatically improved. To this end, we propose a semisupervised learning approach by extending the Gibbs sampling algorithm for modeling context.

ALGORITHM 1: Semisupervised Learning Approach

Input: a sequence of context records R, a set of record-topic tuples R', hyperparameters α, β, γ, the number of context topics K
Global data: count statistics $\{CS\}$, assignment index $\mathcal{A}$
Output: model Θ, Φ, π

Initialization step:

zero all count statistics $\{CS\}$;
for each $\boldsymbol{r_t} \in R$ **do**
  if $\boldsymbol{r_t} \in R'$ **then**
    the context topic assignment $k = c_t$ acc. to $< \boldsymbol{r_t}, c_t > \in R'$;
  else
    sample context topic assignment $k \sim Mult(1/K)$;
  increment statistics of $\{CS\}^k$;
  $\mathcal{A}_t = k$;
end for

Gibbs sampling step:

while not finished **do**
  for each $\boldsymbol{r_t} \in R$ **do**
    the current assignment $k = \mathcal{A}_t$;
    decrement statistics of $\{CS\}^k$;
    if $\boldsymbol{r_t} \in R'$ **then**
      $\hat{k} = k$; //unchanged
    else
      //multinomial sampling acc. to Eq. 5, 15 (B-HMM/HMMC),
      //Eq. 20, 23 (DPMU/DPMUC);
      sample context topic assignment $\hat{k} \sim p(c_t | \boldsymbol{c}_{\neg \boldsymbol{r_t}}, R)$;
    increment statistics of $\{CS\}^{\hat{k}}$;
    $\mathcal{A}_t = \hat{k}$;
  end for
  //check convergence and read out parameters
  if converged or L sampling iterations **then**
    Θ is computed acc. to Eq. 9 (B-HMM/HMMC);
    Φ is computed acc. to Eq. 9 (B-HMM), Eq. 17 (HMMC);
    π is computed acc. to Eq. 17 (HMMC);
    return Θ, Φ, π ;
end while

Specifically, given a set of labeled context records $R' = \{<\boldsymbol{r_1}, c_1>, <\boldsymbol{r_2}, c_2>, \ldots, <\boldsymbol{r_t}, c_t>, \ldots\}$, where $<\boldsymbol{r_t}, c_t>$ is a record-topic tuple, $\boldsymbol{r_t} \in R$, and $c_t \in C$, we denote the K' as the number of unique context topics in R'. We will introduce the details of collecting such labeled data in Section 6.1.1. Then, the objective of our semisupervised learning approach is to integrate R' into the context modeling process and guarantee that the topic assigned to each context record $\boldsymbol{r}$ should be as similar to the tuples in R' as possible. Therefore, an intuitive way is $\forall \boldsymbol{r} \in R'$, where we assign a context topic c to $\boldsymbol{r}$ and guarantee $<\boldsymbol{r}, c> \in R'$. To this end, we propose a semisupervised Gibbs sampling algorithm, and the pseudo-code is shown in Algorithm 1. Specifically, we use $\{CS\}$ to denote the statistics of count (e.g., $n_{\neg \boldsymbol{r_t}, k, j}$) used in the Gibbs sampling process, and $\mathcal{A}$ denotes the assignments of all context records.

In the Initialization step of the algorithm, for each context record $\boldsymbol{r}$, we judge whether $\boldsymbol{r}$ has been labeled. If $\boldsymbol{r}$ has its true label (i.e., $\boldsymbol{r} \in R'$), we directly assign a context topic to $\boldsymbol{r}$ according to $<\boldsymbol{r}, c> \in R'$. Otherwise, we sample a context topic to $\boldsymbol{r}$ from a multinomial distribution $Mult(1/K)$. We then check whether current context record $\boldsymbol{r}$ has a predefined topic label with each iteration of the Gibbs sampling step in R'. If $\boldsymbol{r} \in R'$, we do not assign a context topic to it because we have assigned the context topic in initialization already. Otherwise, as shown in line 9, we assign a context topic to $\boldsymbol{r}$ according to the posterior probability $P(c_t = k|\boldsymbol{c}_{\neg \boldsymbol{r}_t}, R)$, which can be computed refer to Equation (5) (B-HMM), Equation (15) (HMMC), and Equations (20)–(23) (DPMU/DPMUC).

When using this semisupervised algorithm to train DPMU/DPMUC models, in the sampling step (i.e., Step 9), a new context topic can be drawn according to Equation (22). Finally, the parameters are returned after the sampling converges. Particularly for DPMU/DPMUC, the number of unique context topics in $\mathcal{A}$ is the final number of context topics K^*. In addition, since the learning process of DPMU/DPMUC is based on the predefined K' context topics, K^* always satisfies that $K^* \geq K'$.

The inference of proposed models (i.e., B-HMM, HMMC, DPMU, and DPMUC) in this article is derived by the Gibbs sampling method, thus the time complexity is $O(NKT)$, where N is the number of context records, K is the number of context topics, and T is the number of iterations.

6. EXPERIMENTAL RESULTS

In this section, we first introduce the real-world datasets used in our experiments, then we evaluate the efficiency and effectiveness of the proposed approaches for mobile context modeling through extensive experiments. All the experiments are conducted on a 3.10GHz×2 Core CPU, 4GB RAM PC.

6.1. Datasets and Preprocess

Here, we introduce how to collect and preprocess our real-world datasets.

6.1.1. Data Collection. Actually, many mobile apps can help users manage their lives, such as *aTimeLogger,*[1] which allows mobile users to log the time cost of many predefined activities. Illuminated by these Apps, we developed a software named *ActivityLogger* to collect rich context logs from mobile users. *ActivityLogger* provides some predefined activities (e.g., *"Sleep"*) for users, and they can define some activities by themselves. With this software, users can log how long they spend on these activities, thus they can manage their lives more efficiently. Figure 4 shows the screen shots of our software, which mainly has two functions for different tasks, namely *Context Function* and *Activity Function*, respectively. The main task of *Context Function* is to collect rich contextual information from different mobile sensors, such as the system clock, GPS sensor, 3D accelerometers, and the like. Table II illustrates the detailed contextual features collected by our software. Note that the software will automatically start when the mobile device is powered on, and the collection frequency is set to 1 minute. Particularly, users can manually turn off the software if they do not want the current contextual information to be collected due to privacy concerns.

Another function of *ActivityLogger* is the *Activity Function*, which is used to record the starting and ending time of user activities (e.g., when to start and finish working). Table III shows an example of an activity log collected from a mobile user. In our software, there are several predefined activity labels, such as *"Sleep," "Study,"* and *"Entertainment."* Thus, users can choose one label to record their corresponding activities. Particularly, users can also define some new activities by themselves. Intuitively, these activity labels clearly reveal the latent topics of the context records generated by corresponding activities. Each context record $\boldsymbol{r_t}$ contains a timestamp t, and each activity a_i contains a start timestamp $t_{i,1}$ and a end timestamp $t_{i,2}$. Therefore, for each

[1]http://www.atimelogger.com.

Fig. 4. The screen shots of our data collection software *ActivityLogger*.

Table II. The Collected Contextual Features

Data type	Contextual feature	Value range
Time Info	Day Name	{Mon., Tues., Wed., Thur., Fri., Sat., Sun.}
	Holiday	{Yes, No}
	Day Period	{Morning(AM7:00-AM11:00), Noon(AM11:00-PM14:00), Afternoon(PM14:00-PM18:00),Evening(PM18:00-PM21:00), Night(PM21:00-Next day AM7:00)}
	Time Range	{AM0:00-AM1:00,AM1:00-AM2:00,...,PM23:00-PM24:00}
System Info	Profile Type	{General, Silent, Meeting, Outdoor, Pager, Offline}
	Battery Level	{Low(<35%), Middle(35%-65%), High(65%-85%), Full(>85%)}
	Inactive Time	{Short(< 10 minutes), Middle(10-30 minutes), Long(> 30 minutes)}
	Ring Type	{General, Ascending, Ring once, Beep, Silent}.
	Ring Level	{Low(<35%), Middle(35%-75%), High(>75%)}.
GSM Info	Cell ID	Integers.
	Area ID	Integers.
GPS Info	Speed	{Low(< 5km/h), Middle(5–20km/h), High(> 20km/h)}
	Movement	{Yes, No}
	Coordinate	Pair of longitude and latitude.
Event	Applications	{Call, Message, Game, Web browser, Music, Camera, etc.}

context record $\boldsymbol{r_t}$ that satisfies $t_{i,1} \leq t \leq t_{i,2}$, we label its semantic meaning c_t as a_i. Thus, given a sequence of context records $R = \{\boldsymbol{r_1}, \boldsymbol{r_2}, \ldots, \boldsymbol{r_t}, \ldots\}$ and a set of activity records $A = \{a_1, a_2, \ldots, a_i, \ldots\}$, we can easily get a set of labeled data; that is, a set of record-topic tuples $R' = \{<\boldsymbol{r_t}, c_t>\}$.

To collect these context logs, we invited some volunteers with the help of a major mobile device manufacturer and had them install the software *ActivityLogger* on their smartphones. Specifically, we requested all the volunteers to use this software frequently every day and guarantee the authenticity of corresponding activities.

6.1.2. Data Preprocess. To use the activity data for semisupervised context modeling, we need to do some preprocess work to filter out the noise records. Specifically, we first removed those activity records with a duration of less than a threshold ζ_l (we empirically set ζ_l to 5 minutes in our experiments), which may imply that the corresponding

Table III. An Example of an Activity Log

Starting timestamp	Ending timestamp	Activity
2012-12-13-23-21	2012-12-14-08-10	Sleep
2012-12-14-09-10	2012-12-14-11-25	Work
2012-12-14-11-26	2012-12-14-12-00	Eat
2012-12-14-12-01	2012-12-14-13-30	Sleep
2012-12-14-14-00	2012-12-14-15-10	Work
......	......	
2012-12-14-23-50	2012-12-15-09-30	Sleep
2012-12-15-10-21	2012-12-15-11-40	Shopping
2012-12-15-11-41	2012-12-15-13-12	Eat
......	......	
2012-12-22-14-30	2012-12-22-16-00	Work
2012-12-22-16-01	2012-12-22-16-30	Driving
2012-12-22-16-31	2012-12-22-18-00	Gym
2012-12-22-18-21	2012-12-22-19-05	Eat

Table IV. The Details of the 10 Datasets

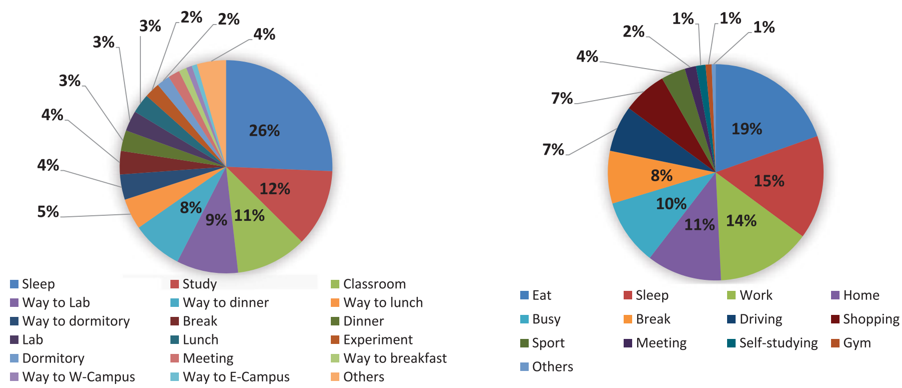
Data set	N	N_L	P	N_p	N_F	$\#L_{Percentage}$
A	115,682	16,215	463	11,232,154	14	14.02%
B	129,021	7,360	702	19,256,150	14	5.70%
C	129,705	3,583	947	24,576,034	16	2.76%
D	129,909	8,652	1,161	33,321,870	16	6.66%
E	129,923	2,488	1,202	36,257,489	16	1.91%
F	129,952	6,459	1,301	42,252,576	16	4.97%
G	130,060	7,968	1,475	51,679,099	16	6.13%
H	130,682	3,329	1,730	66,170,870	16	2.55%
I	131,027	4,654	1,900	79,767,522	16	3.55%
J	131,027	4,264	2,087	88,570,014	16	3.25%

activity record is generated by the wrong operation or just to test the software. Second, we combined some user-defined activities that have similar meanings, such as *“Supermarket”* and *“Shopping.”* Third, some users may start an activity but forget to end it, which may result in a long duration of activity. Thus, to guarantee the accuracy of context labels, we further defined some upper bound thresholds $\{\zeta^a\}$ for each activity. For example, we set ζ^{Eat} to 2 hours, which means that user may not eat something for over 2 hours. Last, the activity records with a frequency of less than 5 are also filtered to guarantee the effectiveness of user labels.

As introduced earlier, in our experiment, we selected the datasets from 10 volunteers spanning 3 months, which are denoted as $\{A, B, \ldots, J\}$. Table IV demonstrates the details of the 10 datasets, where N denotes the number of context records, N_L denotes the number of context records labeled by users through software *ActivityLogger*, P denotes the number of unique contextual feature-value pairs, N_p denotes the occurrence number of all contextual feature-value pairs, N_F denotes the number of contextual features, and $\#L_{Percentage}$ denotes the proportion of labeled data. Moreover, Figure 5 shows the distributions of the labeled context topics of different datasets (here, we only show A and D for conciseness). For each dataset, we use the data from last month as test data and use the remaining data as training data.

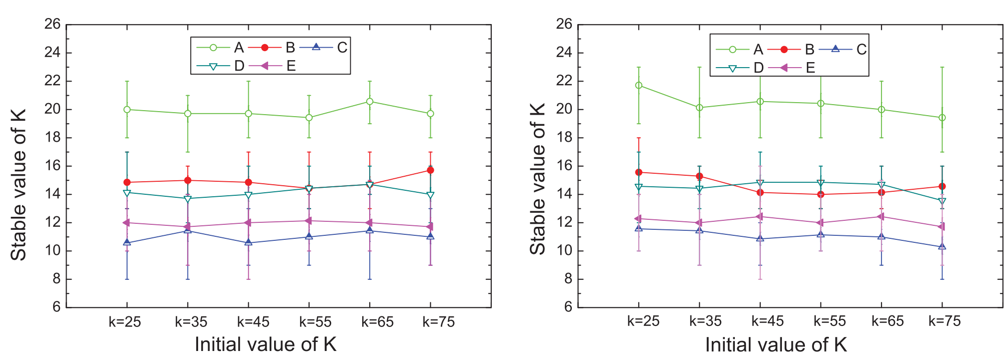
6.2. The Performance of Parameter Estimation by DPMU/DPMUC

In this subsection, we validate the parameter estimation approaches DPMU/DPMUC by evaluating their effectiveness and efficiency.



(a) Dataset A (b) Dataset D

Fig. 5. The distributions of labeled context topics in datasets A and D.

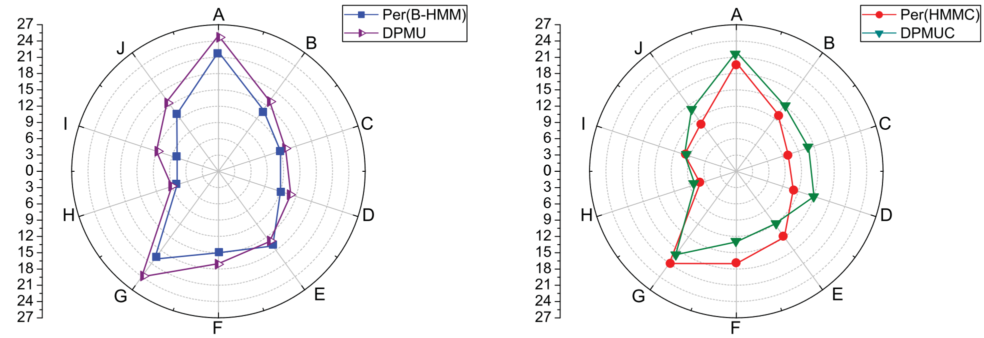


(a) DPMU (b) DPMUC

Fig. 6. The results of DPMU and DPMUC with different initial values of K.

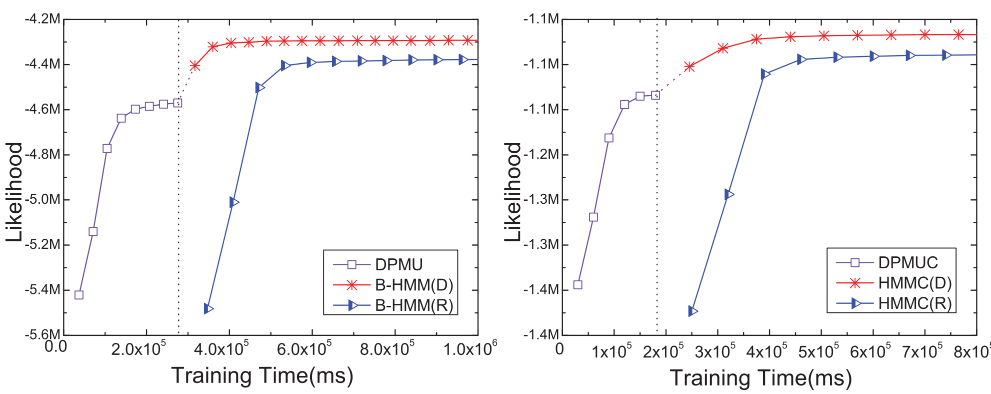
6.2.1. Effectiveness of the DPMU/DPMUC. To estimate the number of context topics, we propose the DPMU/DPMUC approaches for automatically learning the best number of topics when given only an initial value. Here, we validate the effectiveness of DPMU/DPMUC. First, we evaluate whether DPMU/DPMUC can get a stable result when given different initial values. Specifically, we set the initial values of the number of context topics from $\mathbb{K} = \{25, 35, 45, 55, 65, 75\}$, and for each $K_i \in \mathbb{K}$, we run DPMU/DPMUC 10 times. Figure 6 shows the box chart of the experimental results (since the results of the 10 datasets have similar trends, we only show the results of the first 5 datasets $\{A, \ldots, E\}$). Note that the convergence of the Gibbs sampling processes of DPMU/DPMUC are measured by the log likelihood of the training data, and the parameters α and β are empirically set to 100 and 10, respectively. From this figure, we can observe that DPMU/DPMUC can obtain stable results with different initial values.

Second, we check whether the number K obtained by DPMU/DPMUC is reasonable. To this end, we first choose the commonly used perplexity-based approach [Bao et al. 2012] to estimate the number of latent context topics for B-HMM and HMMC models as a benchmark. Then, we compare the difference between the results from the perplexity-based approach and our DPMU/DPMUC model. Figure 7 shows the number of context topics learned by the perplexity-based approach and DPMU/DPMUC for B-HMM and



(a) DPMU (b) DPMUC

Fig. 7. Spherical comparison in terms of the number of context topics.



(a) DPMU (b) DPMUC

Fig. 8. The comparison in terms of the likelihood curves of dataset A.

HMMC on the 10 datasets. From the results, we can observe that the number K learned from the DPMU/DPMUC model is very close to that learned from the perplexity-based approach, which validates the effectiveness of our DPMU/DPMUC approaches. Note that here the value K learned by DPMU/DPMUC is the average value of results shown in Figure 6.

6.2.2. Efficiency of the DPMU/DPMUC. As discussed in Section 4.3, the training results of DPMU/DPMUC can be used to train B-HMM and HMMC models. Indeed, the basic Gibbs sampling method is based on the random assignment of initial values, which may result in the high computational cost (i.e., iterations), and local optimal values. At the same time, after the Gibbs sampling of DPMU/DUPMC, we can get a set of the topic assignments for context records. Therefore, we can use these assignments as the initial values for training B-HMM and HMMC models. Figure 8 shows the convergence curves of Gibbs sampling for B-HMM and HMMC models with different initial value assignment by measuring their log likelihood for the dataset A (the convergence curves for other training datasets follow the similar trend). Note that the left end of the dotted line in Figure 8 is the curve of the likelihood of training DPMU and DPMUC. B-HMM(D)/HMMC(D) denotes using initial value assignment through DPMU/DPMUC models, and B-HMM(R)/HMMC(R) denotes using random initial value assignment.

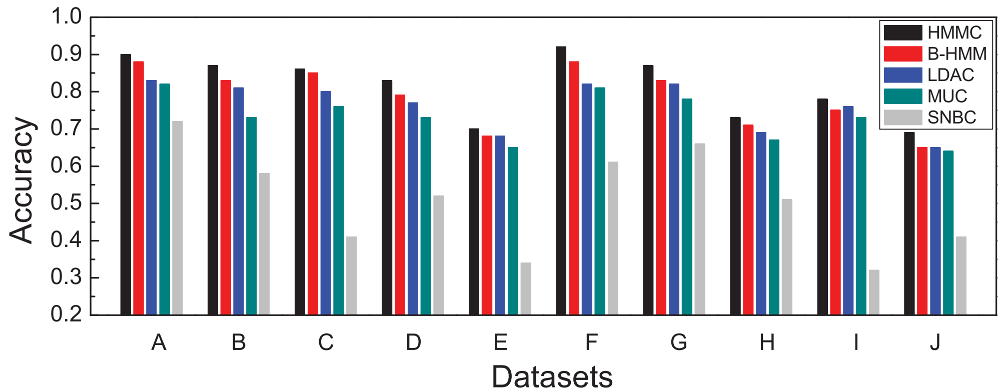


Fig. 9. The comparison of the label *Accuracy* between different approaches.

From this figure, we can observe that assigning initial values through DPMU/DPMUC model can accelerate the training processes of both B-HMM and HMMC models. Moreover, we also find that the B-HMM and HMMC models with initial value assignment by DPMU/DPMUC can achieve higher optimal points, which indicates the higher likelihood of model training.

6.3. The Performance of B-HMM and HMMC Models

In this subsection, we validate our B-HMM and HMMC models by evaluating their performance on context recognition.

6.3.1. The Performance of Labeling Context Topics. Many user activity labels can be used for semisupervised learning in B-HMM and HMMC models. Here, we first evaluate whether B-HMM and HMMC models can be used to label context topics (user activity labels) in test datasets. To the best of our knowledge, there is no existing semisupervised approach for modeling context. Thus, we extend two state-of-the-art unsupervised approaches, LDAC and MUC [Bao et al. 2012], as baselines to evaluate our approaches. To be specific, we use the semisupervised Gibbs sampling approach introduced in Algorithm 1 to extend LDAC and MUC to label context topics. Moveover, we also develop a sequence approach, Sequence-based Naive Bayes on Context (SNBC), for evaluating our approaches, which can label context records by

$$
\begin{aligned}
P(c_t = k|\boldsymbol{r_t}) &\propto P(\boldsymbol{r_t}|c_t = k)P(c_t|c_{t-1}) \\
&\propto P(c_t = k|c_{t-1}) \prod_{i=1}^{N_t} P(r_{t,i}|c_t = k),
\end{aligned}
$$

where $P(c_t = k|c_{t-1}) = \frac{1}{K}$ for the context record $\boldsymbol{r_t}$ without previous record $\boldsymbol{r}_{t-1}$.

In our experiments, we use the user activity labels in test dataset as ground truth, and we test the *Accuracy* of context labels predicted by each approach. Specifically, we denote N_T as the number of times that a context record is labeled with right topic, while N_F is the number of times that a context record is labeled with wrong topic; thus, the accuracy is calculated by $Accuracy = \frac{N_T}{N_T+N_F}$. Figure 9 shows the results of the comparison of label accuracy between different approaches in 10 datasets. From this figure, we can observe that LDAC and MUC have competitive performance in label accuracy, and our B-HMM and HMMC models have the best label accuracy.

Furthermore, to validate the performance of context recognition with respect to different context topics, we also compute the overall *Precision*, *Recall*, and F_1 score for

Table V. The Comparison of the Label *Precision* between Different Approaches

Model	A	B	C	D	E	F	G	H	I	J
HMMC	**0.900**	**0.873**	**0.862**	**0.821**	**0.702**	**0.921**	**0.831**	**0.729**	**0.782**	**0.721**
B-HMM	0.862	0.832	0.843	0.785	0.681	0.879	0.828	0.713	0.746	0.702
LDAC	0.821	0.817	0.832	0.776	0.682	0.829	0.822	0.695	0.767	0.688
MUC	0.793	0.738	0.765	0.732	0.650	0.811	0.782	0.679	0.731	0.671
SNBC	0.721	0.582	0.413	0.520	0.343	0.619	0.668	0.516	0.329	0.392

Table VI. The Comparison of the Label *Recall* between Different Approaches

Model	A	B	C	D	E	F	G	H	I	J
HMMC	**0.821**	**0.796**	**0.736**	**0.791**	**0.679**	**0.911**	**0.723**	**0.701**	**0.751**	**0.656**
B-HMM	0.813	0.782	0.719	0.772	0.662	0.855	0.712	0.698	0.733	0.622
LDAC	0.705	0.721	0.688	0.718	0.602	0.812	0.658	0.667	0.678	0.602
MUC	0.691	0.703	0.671	0.719	0.592	0.789	0.669	0.651	0.682	0.611
SNBC	0.530	0.411	0.422	0.451	0.302	0.618	0.490	0.529	0.450	0.412

Table VII. The Comparison of the Label F_1 Score between Different Approaches

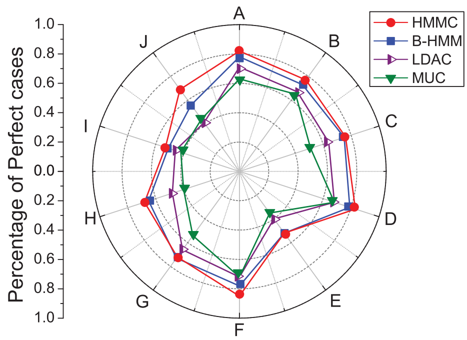
Model	A	B	C	D	E	F	G	H	I	J
HMMC	**0.859**	**0.833**	**0.794**	**0.806**	**0.690**	**0.916**	**0.773**	**0.715**	**0.766**	**0.687**
B-HMM	0.837	0.806	0.776	0.778	0.671	0.867	0.766	0.705	0.739	0.660
LDAC	0.759	0.766	0.753	0.746	0.640	0.820	0.731	0.681	0.720	0.642
MUC	0.738	0.720	0.715	0.725	0.620	0.800	0.721	0.665	0.706	0.640
SNBC	0.611	0.482	0.417	0.483	0.321	0.618	0.565	0.522	0.380	0.402

each approach in our datasets. Specifically, we first compute the $Precision$, $Recall$, and F_1 score for each context topic in the dataset, and then get the average value across all context topics as the overall performance. If we denote N_T^c as the number of times that a context record is labeled as c correctly, N_F^c as the number of times that a context record is labeled as c incorrectly, and N^c as the number of context records with label c truly, we can compute the $Precision$, $Recall$, and F_1 score of a given context topic c by $P_c = \frac{N_T^c}{N_T^c + N_F^c}$, $R_c = \frac{N_T^c}{N^c}$, and $F_1 = \frac{2*P_c*R_c}{P_c+R_C}$. Tables V, VI, and VII show the results of the performance of different approaches with respect to different evaluation metrics. From these tables, we can observe that our approaches consistently outperform other baselines with respect to all metrics.

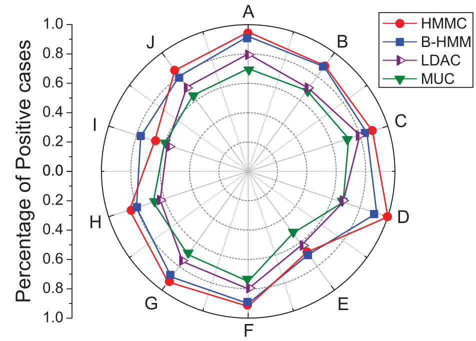
From these experimental results, we come to several conclusions. First, sequence dependence is an important characteristic of contextual information, and considering sequence dependence when context modeling can enhance the performance of context recognition. Second, only using the characteristic of sequence dependence for context recognition is not enough (e.g., SNBC has the worst recognition performance) since it cannot capture the latent semantics of contextual information. Last, the prior knowledge of contextual features can be used for modeling contextual information more accurately (i.e., HMMC outperforms B-HMM slightly).

6.3.2. Human Judgement-based Evaluation. To evaluate the quality of learned context topics, in the experiments, we also asked the 10 volunteers in our datasets to manually judge the context topics learned from their own context logs. Specifically, for each learned context topic c_k, we chose the contextual feature-value pairs p with $P(p|c_k) > 0.01$ to represent the context c_k and showed them to corresponding volunteers for judgment.

Similar to the user study introduced in Bao et al. [2012], we also prepared three remarks for volunteers use in judging the quality of each learned context:



(a)



(b)

Fig. 10. Spherical comparison in terms of human evaluation.

—P: ***P**erfect*. This remark indicates that the learned context topic reflects one of the volunteer's typical contexts well. No irrelevant context information is included, and no relevant context information is missing.
—G: ***G**ood*. This remark indicates that the learned context topic partially reflects one of the volunteer's typical contexts but contains some irrelevant context information or misses some relevant information.
—B: ***B**ad*. This remark indicates that it is hard to state the learned context that reflects the typical context topics of a volunteer.

We denote the cases with perfect or good remarks as positive cases. Figure 10 shows the results of human judgement for each approach. From this figure, we can observe that both B-HMM and HMMC models outperform two unsupervised models (i.e., LDAC and MUC), which indicates that the context topics learned by our models are more reasonable because our models can integrate both sequential characteristics and user labels of mobile contexts into the learning process.

6.4. Case Studies for Evaluating B-HMM and HMMC Models

Here, we study two cases for evaluating our B-HMM and HMMC models.

6.4.1. Case Study 1. To evaluate the quality of learned context topics more intuitively, we studied some typical context topics learned by each approach. Specifically, we first contacted one volunteer from dataset A; we knew he is a postgraduate with a typical personalized habit of *"usually exercising at gym on Friday and Saturday afternoon (16:00-18:00)."* Then, we manually checked the latent context topics learned by MUC, LDAC, B-HMM, and HMMC and interestingly found that all of these approaches can discover a context topic that is relevant to this typical habit. Tables VIII–XI show the related context topics learned by MUC, LDAC, B-HMM, and HMMC, respectively, which are denoted as c_m, c_l, c_b, and c_h for simplicity. In addition, the contextual feature *Location* has been manually translated to meaningful locations to help Understanding.

From these figures, we find that most of the contextual feature-value pairs in c_m and c_l are reasonable, such as *(Day Name: Saturday)*, *(Location:gym)* and *(Day Period: Afternoon)*. However, there are also some irrelevant contextual feature-values pairs in c_m and c_l, which are shown in bold. Moreover, both of the two context topics miss the important contextual feature-value pair *(Day Name: Friday)*. By contrast, we find most of the contextual feature-value pairs in c_b and c_h are reasonable, except for two irrelevant pairs in c_b. As a result, both c_m and c_l are labeled as *"Bad,"* c_b is labeled as *"Good,"* and c_h is labeled as *"Perfect."*

Table VIII. Context c_m Learned by MUC

(Profile Type: General)
(Day Name: Saturday)
(Area ID: 21885)
(Cell ID: id_1)
(Location: gym)
(Day Period: Noon)
(Day Period: Afternoon)
(Time Range: PM11:00-12:00)
(Time Range: PM12:00-13:00)
(Time Range: PM16:00-17:00)
(Movement: No)
(Battery Level: High(65%-85%))
(Inactive Time: Long(>30 minutes))

Table IX. Context c_l Learned by LDAC

(Profile Type: General)
(Day Name: Saturday)
(Area ID: 21885)
(Cell ID: id_1)
(Location: gym)
(Day Period: Noon)
(Day Period: Afternoon)
(Time Range: PM12:00-13:00)
(Time Range: PM16:00-17:00)
(Time Range: PM17:00-18:00)
(Movement: No)
(Battery Level: Middle(35%-65%))
(Inactive Time: Long(>30 minutes))

Table X. Context c_b Learned by B-HMM

(Profile Type: General)
(Day Name: Saturday)
(Day Name: Friday)
(Area ID: 21885)
(Cell ID: id_1)
(Location: gym)
(Day Period: Noon)
(Day Period: Afternoon)
(Time Range: PM12:00-13:00)
(Time Range: PM16:00-17:00)
(Time Range: PM17:00-18:00)
(Movement: No)
(Battery Level: High(65%-85%))
(Battery Level: Middle(35%-65%))

Table XI. Context c_h Learned by HMMC

(Profile Type: General)
(Day Name: Saturday)
(Day Name: Friday)
(Area ID: 21885)
(Cell ID: id_1)
(Location: gym)
(Time Range: PM16:00-17:00)
(Time Range: PM17:00-18:00)
(Movement: No)
(Battery Level: High(65%-85%))
(Battery Level: Middle(35%-65%))
(Inactive Time: Long(>30 minutes))

6.4.2. Case Study 2. Our B-HMM and HMMC models can consider the sequence dependence of context records. Therefore, to further check the characteristics of sequence dependence, we illustrate part of the context topic transitions based on Θ learned by the B-HMM model in dataset A and by the HMMC model in dataset D.

For the conciseness, here we only show five typical context topics and the top five context topics with the highest transition probability from them in Tables XIII and XII. Note that some of the context topics in the tables are labeled by volunteers from the datasets, and others are labeled manually by ourselves. From these tables, we find that all these transitions are reasonable and easy to understand. For example, the volunteer from dataset A is a postgraduate student; thus, he has some typical context transition patterns like "Way to Lab $\longrightarrow$ Lab $\longrightarrow$ Experiments.' Also, the volunteer from dataset D is a white-collar worker in a software company; thus, he has some typical context transition patterns like "Work $\longrightarrow$ Driving $\longrightarrow$ Eat." Moreover, we find that for each context topic c_i, the next context topic with highest transition probabilities is c_i itself. This is reasonable because the activities of mobile users usually last for a while; thus, the probability that the next context topic is similar to the current context is high. Indeed, in our B-HMM and HMMC models, the context topic for each context record is not only decided by the transition probability Θ, but also by the emission probability Φ and contextual feature distribution π.

Table XII. Context Topic Transitions Learned by B-HMM from Dataset *A*

	top1	top2	top3	top4	top5
Dormitory	Dormitory	Sleep	Way to lab	Way to lunch	Classroom
Study	Study	Way to lunch	Way to dinner	Way to dormitory	Break
Way to Lab	Way to Lab	Lab	Experiment	Study	Break
Sleep	Sleep	Dormitory	Way to breakfast	Way to lunch	Break
Meeting	Meeting	Way to dinner	Way to dormitory	Study	Break

Table XIII. Context Topic Transitions Learned by HMMC from Dataset *D*

	top1	top2	top3	top4	top5
Work	Work	Driving	Eat	Sport	Meeting
Sleep	Sleep	Home	Eat	Self-studying	TV
Gym	Gym	Eat	Break	Shopping	Driving
Shopping	Shopping	Eat	Driving	Home	Break
Eat	Eat	Home	Work	Break	Shopping

7. CONCLUDING REMARKS

In this article, we provided a semisupervised approach to modeling the context of mobile users. Specifically, we first exploited the B-HMM for modeling user contexts in the form of probabilistic distributions and transitions of raw context data. Also, we developed a sequential model by extending B-HMM with the prior knowledge of contextual features. In this way, we are able to model context more accurately. Next, to efficiently learn the parameters and initial values of the proposed models, we developed a novel approach for parameter estimation by integrating the DPM and MU model. Furthermore, by incorporating both user-labeled and unlabeled data, we proposed a straightforward but efficient semisupervised learning-based framework to model and discover context topics. Finally, experimental results on real-world datasets clearly validated both the efficiency and effectiveness of the proposed approaches on recognizing the personalized context of mobile users.

In the future, we would like to investigate whether and how other types of information, such as user demographics, can help to improve the performance of context modeling. Furthermore, it will also be interesting to investigate how to protect the privacy of mobile users during context modeling.

APPENDIX A. INFERENCE AND PARAMETER ESTIMATION

In this appendix, we introduce the Gibbs sampling derivation of the equations used in this paper. To make it simple for understanding, we just show the derivation of some typical equations.

A.1. Inference $P(c_t|c_{\neg r_t}, R)$ of B-HMM

$$\begin{aligned}
P(c_t|\boldsymbol{c}_{\neg \boldsymbol{r}_t}, R) &\propto P(c_t, \boldsymbol{c}_{\neg \boldsymbol{r}_t}, R) \\
&= P(R|c_t, \boldsymbol{c}_{\neg \boldsymbol{r}_t})P(c_t|\boldsymbol{c}_{\neg \boldsymbol{r}_t})P(\boldsymbol{c}_{\neg \boldsymbol{r}_t}) \\
&= P(R|c_t, \boldsymbol{c}_{\neg \boldsymbol{r}_t})P(c_t|\boldsymbol{c}_{\neg \boldsymbol{r}_t})P(c_{t+1}, \boldsymbol{c}_{\neg(\boldsymbol{r}_t, \boldsymbol{r}_{t+1})}) \\
&= P(R|c_t, \boldsymbol{c}_{\neg \boldsymbol{r}_t})P(c_t|\boldsymbol{c}_{\neg \boldsymbol{r}_t})P(c_{t+1}|\boldsymbol{c}_{\neg(\boldsymbol{r}_t, \boldsymbol{r}_{t+1})})P(\boldsymbol{c}_{\neg(\boldsymbol{r}_t, \boldsymbol{r}_{t+1})}) \\
&\propto P(R|c_t, \boldsymbol{c}_{\neg \boldsymbol{r}_t})P(c_t|\boldsymbol{c}_{\neg \boldsymbol{r}_t})P(c_{t+1}|\boldsymbol{c}_{\neg(\boldsymbol{r}_t, \boldsymbol{r}_{t+1})}) \\
&= P(\boldsymbol{r_t}|c_t, \boldsymbol{c}_{\neg \boldsymbol{r}_t}, \Lambda_{\neg \boldsymbol{r}_t})P(\Lambda_{\neg \boldsymbol{r}_t}|c_t, \boldsymbol{c}_{\neg \boldsymbol{r}_t})P(c_t|\boldsymbol{c}_{\neg \boldsymbol{r}_t})P(c_{t+1}|\boldsymbol{c}_{\neg(\boldsymbol{r}_t, \boldsymbol{r}_{t+1})}) \\
&= P(\boldsymbol{r_t}|c_t, \boldsymbol{c}_{\neg \boldsymbol{r}_t}, \Lambda_{\neg \boldsymbol{r}_t})P(\Lambda_{\neg \boldsymbol{r}_t}|\boldsymbol{c}_{\neg \boldsymbol{r}_t})P(c_t|\boldsymbol{c}_{\neg \boldsymbol{r}_t})P(c_{t+1}|\boldsymbol{c}_{\neg(\boldsymbol{r}_t, \boldsymbol{r}_{t+1})}) \\
&\propto P(\boldsymbol{r_t}|c_t, \boldsymbol{c}_{\neg \boldsymbol{r}_t}, \Lambda_{\neg \boldsymbol{r}_t})P(c_t|\boldsymbol{c}_{\neg \boldsymbol{r}_t})P(c_{t+1}|\boldsymbol{c}_{\neg(\boldsymbol{r}_t, \boldsymbol{r}_{t+1})}).
\end{aligned} \tag{25}$$

In Appendix A.3 and A.4, we will show the detailed Gibbs sampling derivation of $P(\boldsymbol{p_t}|c_t, \boldsymbol{c}_{\neg\boldsymbol{p_t}}, \Lambda_{\neg\boldsymbol{p_t}})$ and $P(c_t|\boldsymbol{c}_{\neg\boldsymbol{p_t}})$.

A.2. Inference $P(c_t|c_{\neg r_t}, R)$ of HMMC

$$\begin{aligned} P(c_t|\boldsymbol{c}_{\neg\boldsymbol{r_t}}, R) &\propto P(c_t, \boldsymbol{c}_{\neg\boldsymbol{r_t}}, R) \\ &\propto P(c_t, \boldsymbol{c}_{\neg\boldsymbol{r_t}}, V, F) \\ &\propto P(V|c_t, \boldsymbol{c}_{\neg\boldsymbol{r_t}}, F)P(c_t|\boldsymbol{c}_{\neg\boldsymbol{r_t}})P(\boldsymbol{c}_{\neg\boldsymbol{r_t}})P(F) \\ &\propto P(\boldsymbol{v_t}|c_t, \boldsymbol{c}_{\neg\boldsymbol{r_t}}, F, V_{\neg\boldsymbol{r_t}})P(c_t|\boldsymbol{c}_{\neg\boldsymbol{r_t}})P(c_{t+1}|\boldsymbol{c}_{\neg(\boldsymbol{r_t},\boldsymbol{r_{t+1}})}). \end{aligned} \tag{26}$$

A.3. Inference $P(\boldsymbol{r_t}|\boldsymbol{c_t}, \boldsymbol{c}_{\neg r_t}, \Lambda_{\neg r_t})$

$$P(\boldsymbol{r_t}|c_t, \boldsymbol{c}_{\neg\boldsymbol{r_t}}, \Lambda_{\neg\boldsymbol{r_t}}) \propto P(\boldsymbol{r_t}, \boldsymbol{c}|\beta) \tag{27}$$

$$= \int P(\boldsymbol{r_t}|\boldsymbol{c}, \Phi)P(\Phi|\beta)d\Phi. \tag{28}$$

To compute this value by integrating out the multinomial parameter Φ from the probability, we have

$$= \int \prod_{k=1}^{K} P(\phi_k|\beta) \prod_{t=1}^{|R|} \prod_{i=1}^{N_t} P(p_{t,i}|c_t, \phi_{c_t})d\phi. \tag{29}$$

Then, we distribute the integral over the probability per context topic k,

$$= \prod_{k=1}^{K} \int P(\phi_k|\beta) \prod_{t=1}^{|R|} \prod_{i=1}^{N_t} P(p_{t,i}|c_t, \phi_{c_t})d\phi_k. \tag{30}$$

Next, we expand the probability formula based on its density (i.e., Dirichlet and Multinomial),

$$= \prod_{k=1}^{K} \int \frac{\Gamma\left(\sum_{j=1}^{J} \beta_j\right)}{\prod_{j=1}^{J} \Gamma(\beta_j)} \prod_{j=1}^{J} \phi_{k,j}^{\beta_j - 1} \prod_{t=1}^{|R|} \prod_{i=1}^{N_t} \phi_{c_t, p_{t,i}} d\phi_k, \tag{31}$$

where, J is the number of unique contextual feature-value pairs, $|R|$ is the number of all context records, N_t is the number of contextual feature-value pairs in the context record (i.e., $\boldsymbol{r_t}$) whose timestamp is t, and $p_{t,i}$ is the i-th contextual feature-value pair in record $\boldsymbol{r_t}$. Based on the equality relationship $\prod_{t=1}^{|R|} \prod_{i=1}^{N_t} \phi_{c_t, p_{t,i}} = \prod_{j=1}^{J} \phi_{k,j}^{n_{k,j}}$, we have,

$$= \prod_{k=1}^{K} \int \frac{\Gamma\left(\sum_{j=1}^{J} \beta_j\right)}{\prod_{j=1}^{J} \Gamma(\beta_j)} \prod_{j=1}^{J} \phi_{k,j}^{\beta_j - 1} \prod_{j=1}^{J} \phi_{k,j}^{n_{k,j}} d\phi_k. \tag{32}$$

Since $x^a x^b = x^{a+b}$,

$$= \prod_{k=1}^{K} \int \frac{\Gamma\left(\sum_{j=1}^{J} \beta_j\right)}{\prod_{j=1}^{J} \Gamma(\beta_j)} \prod_{j=1}^{J} \phi_{k,j}^{n_{k,j}+\beta_j - 1} d\phi_k. \tag{33}$$

Then, we multiply by 1 expressed in the following convenient from and distribute through the integral,

$$= \prod_{k=1}^{K} \frac{\Gamma\left(\sum_{j=1}^{J} \beta_j\right)}{\prod_{j=1}^{J} \Gamma(\beta_j)} \frac{\prod_{j=1}^{J} \Gamma(n_{k,j} + \beta_j)}{\Gamma\left(\sum_{j=1}^{J} n_{k,j} + \beta_j\right)} \int \frac{\Gamma\left(\sum_{j=1}^{J} n_{k,j} + \beta_j\right)}{\prod_{j=1}^{J} \Gamma(n_{k,j} + \beta_j)} \prod_{j=1}^{J} \phi_{k,j}^{n_{k,j}+\beta_j - 1} d\phi_k. \tag{34}$$

Because the remaining integral is of the Dirichlet density over its complete support (thus the value is equal to 1), we can drop out of the product. Also, we can remove the constant terms.

$$\propto \prod_{k=1}^{K} \frac{\prod_{j=1}^{J} \Gamma(n_{k,j} + \beta_j)}{\Gamma\left(\sum_{j=1}^{J} n_{k,j} + \beta_j\right)}. \tag{35}$$

Now we can pull apart the equation based on whether the context topic k is the assignment c_t to the current context record $\boldsymbol{r_t}$,

$$= \prod_{k \neq c_t} \frac{\prod_{j=1}^{J} \Gamma\left(n_{\neg \boldsymbol{r_t},k,j} + \beta_j\right)}{\Gamma\left(\sum_{j=1}^{J} n_{\neg \boldsymbol{r_t},k,j} + \beta_j\right)} \times \frac{\prod_{j=1}^{J} \Gamma\left(n_{\neg \boldsymbol{r_t},k,j} + N_{t,j} + \beta_j\right)}{\Gamma\left(\sum_{j=1}^{J} n_{\neg \boldsymbol{r_t},k,j} + N_{t,j} + \beta_j\right)}. \tag{36}$$

Based on the expansion $\Gamma(x+1) = x \times \Gamma(x)$, we can expand out the increment term,

$$= \prod_{k \neq c_t} \frac{\prod_{j=1}^{J} \Gamma(n_{\neg \boldsymbol{r_t},k,j} + \beta_j)}{\Gamma(\sum_{j=1}^{J} n_{\neg \boldsymbol{r_t},k,j} + \beta_j)} \times \frac{\prod_{j=1}^{J} (\Gamma(n_{\neg \boldsymbol{r_t},c_t,j} + \beta_j) \times \prod_{i=1}^{N_{t,j}} (n_{\neg \boldsymbol{r_t},c_t,j} + \beta_j + i))}{\Gamma(\sum_{j=1}^{J} n_{\neg \boldsymbol{r_t},c_t,j} + \beta_j) \times \prod_{i=1}^{N_t} (\sum_{j=1}^{J} n_{\neg \boldsymbol{r_t},c_t,j} + \beta_j + i)}. \tag{37}$$

Then, we refold the residual Γ-function terms back into their products,

$$= \prod_{k=1}^{K} \frac{\prod_{j=1}^{J} \Gamma(n_{\neg \boldsymbol{r_t},k,j} + \beta_j)}{\Gamma\left(\sum_{j=1}^{J} n_{\neg \boldsymbol{r_t},k,j} + \beta_j\right)} \times \frac{\prod_{j=1}^{J} \prod_{i=1}^{N_{t,j}} (n_{\neg \boldsymbol{r_t},c_t,j} + \beta_j + i)}{\prod_{i=1}^{N_t} \left(\sum_{j=1}^{J} n_{\neg \boldsymbol{r_t},c_t,j} + \beta_j + i\right)}. \tag{38}$$

Finally, we eliminate the terms which do not depend on c_t,

$$= \frac{\prod_{j=1}^{J} \prod_{i=1}^{N_{t,j}} (n_{\neg \boldsymbol{r_t},c_t,j} + \beta_j + i)}{\prod_{i=1}^{N_t} (\sum_{j=1}^{J} n_{\neg \boldsymbol{r_t},c_t,j} + \beta_j + i)}. \tag{39}$$

A.4. Inference $P(c_t | c_{\neg r_t})$

It should be noted that, Θ is the context topic transition distribution (a $K \times K$ matrix),

$$P(c_t | \boldsymbol{c}_{\neg \boldsymbol{r_t}}) \propto \int \prod_{k=1}^{K} P(\theta_k | \alpha) \prod_{t=1}^{|R|} P(c_t | \theta_k) d\theta \tag{40}$$

$$= \prod_{k=1}^{K} \int P(\theta_k | \alpha) \prod_{t=1}^{|R|} P(c_t | \theta_k) d\theta_k. \tag{41}$$

Expand the probability formula based on the corresponding density,

$$= \prod_{k=1}^{K} \int \frac{\Gamma\left(\sum_{k'=1}^{K} \alpha_{k'}\right)}{\prod_{k'=1}^{K} \Gamma(\alpha_{k'})} \prod_{k'=1}^{K} \theta_{k,k'}^{\alpha_{k'}-1} \prod_{k'=1}^{K} \theta_{k,k'}^{n_{k,k'}} d\theta_k, \tag{42}$$

where $n_{k,k'}$ denotes the number of context records assigned by k' and the previous record is assigned by k. Based on $x^a x^b = x^{a+b}$, we get,

$$= \prod_{k=1}^{K} \int \frac{\Gamma\left(\sum_{k'=1}^{K} \alpha_{k'}\right)}{\prod_{k'=1}^{K} \Gamma(\alpha_{k'})} \prod_{k'=1}^{K} \theta_{k,k'}^{n_{k,k'}+\alpha_{k'}-1} d\theta_k. \tag{43}$$

Similar to A.3, we try to add some formulas to construct a convenient form,

$$= \prod_{k=1}^{K} \frac{\Gamma\left(\sum_{k'=1}^{K} \alpha_{k'}\right)}{\prod_{k'=1}^{K} \Gamma(\alpha_{k'})} \frac{\prod_{k'=1}^{K} \Gamma(n_{k,k'} + \alpha_{k'})}{\Gamma\left(\sum_{k'=1}^{K} n_{k,k'} + \alpha_{k'}\right)} \int \frac{\Gamma\left(\sum_{k'=1}^{K} n_{k,k'} + \alpha_{\alpha_{k'}}\right)}{\prod_{k'=1}^{K} \Gamma(n_{k,k'} + \alpha_{k'})} \prod_{k'=1}^{K} \theta_{k,k'}^{n_{k,k'}+\alpha_{k'}-1} d\theta_k. \tag{44}$$

Remove all the constant terms and that do not depend on the current context topic c_t,

$$\propto \prod_{k=1}^{K} \frac{\prod_{k'=1}^{K} \Gamma(n_{k,k'} + \alpha_{k'})}{\Gamma\left(\sum_{k'=1}^{K} n_{k,k'} + \alpha_{\alpha_{k'}}\right)}. \tag{45}$$

Then it is time to pull apart the products based on whether the context topic k is the assignment c_t to $\boldsymbol{r_t}$ with timestamp t. Note that, one context record only assigned once in each iteration, thus the value is only need plus 1 count, then we get,

$$= \prod_{k=1}^{K} \frac{\prod_{k' \neq c_t} \Gamma(n_{\neg \boldsymbol{r_t},(k,k')} + \alpha_{k'})}{\Gamma\left(1 + \sum_{k'=1}^{K} n_{\neg \boldsymbol{r_t},(k,k')} + \alpha_{k'}\right)} \Gamma(n_{\neg \boldsymbol{r_t},(k,c_t)} + \alpha_{c_t} + 1), \tag{46}$$

Recall that k in the equation denotes the context topic of the previous context record. In our problem, the previous context topic of $\boldsymbol{r_t}$ is given, and denoted as c_{t-1}. Thus, we can remove $\prod_{k=1}^{K}$ and replace k by c_{t-1},

$$= \frac{\prod_{k' \neq c_t} \Gamma(n_{\neg \boldsymbol{r_t},(c_{t-1},k')} + \alpha_{k'})}{\Gamma\left(1 + \sum_{k'=1}^{K} n_{\neg \boldsymbol{r_t},(c_{t-1},k')} + \alpha_{k'}\right)} \Gamma(n_{\neg \boldsymbol{r_t},(c_{t-1},c_t)} + \alpha_{c_t}) \times (n_{\neg \boldsymbol{r_t},(c_{t-1},c_t)} + \alpha_{c_t}). \tag{47}$$

Refold the residual Γ-function terms back into the general products,

$$= \frac{\prod_{k'=1}^{K} \Gamma(n_{\neg \boldsymbol{r_t},(c_{t-1},k')} + \alpha_{k'})}{\Gamma\left(1 + \sum_{k'=1}^{K} n_{\neg \boldsymbol{r_t},(c_{t-1},k')} + \alpha_{k'}\right)} \times (n_{\neg \boldsymbol{r_t},(c_{t-1},c_t)} + \alpha_{c_t}). \tag{48}$$

Transform the denominator based on $\Gamma(x+1) = x \times \Gamma(x)$,

$$= \frac{\prod_{k'=1}^{K} \Gamma(n_{\neg \boldsymbol{r_t},(c_{t-1},k')} + \alpha_{k'})}{\Gamma\left(\sum_{k'=1}^{K} n_{\neg \boldsymbol{r_t},(c_{t-1},k')} + \alpha_{k'}\right) \times \left(\sum_{k'=1}^{K} n_{\neg \boldsymbol{r_t},(c_{t-1},k')} + \alpha_{k'}\right)} \times (n_{\neg \boldsymbol{r_t},(c_{t-1},c_t)} + \alpha_{c_t}) \tag{49}$$

$$\propto \frac{n_{\neg \boldsymbol{r_t},(c_{t-1},c_t)} + \alpha_{c_t}}{\sum_{k'=1}^{K} n_{\neg \boldsymbol{r_t},(c_{t-1},k')} + \alpha_{k'}} \tag{50}$$

$$= \frac{n_{\neg \boldsymbol{r_t},(c_{t-1},c_t)} + \alpha_{c_t}}{n_{\neg \boldsymbol{r_t},(c_{t-1},*)} + \sum_{k'=1}^{K} \alpha_{k'}}. \tag{51}$$

REFERENCES

G. D. Abowd, C. G. Atkeson, J. Hong, S. Long, R. Kooper, and M. Pinkerton. 1997. Cyberguide: A mobile context-aware tour guide. *Wireless Networks* 3, 5, 421–433.

J. H. Albert and S. Chib. 1993. Bayes inference via Gibbs sampling of autoregressive time series subject to Markov mean and variance shifts. *Journal of Business & Economic Statistics* 11, 1, 1–15.

C. B. Anagnostopoulos, A. Tsounis, and S. Hadjiefthymiades. 2007. Context awareness in mobile computing environments. *Wireless Personal Communications* 42, 3, 445–464.

C. E. Antoniak. 1974. Mixtures of dirichlet processes with applications to Bayesian nonparametric problems. *Annals of Statistics* 1152–1174.

L. Azzopardi, M. Girolami, and K. Van Risjbergen. 2003. Investigating the relationship between language model perplexity and IR precision-recall measures. In *Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*. ACM, 369–370.

T. Bao, H. Cao, E. Chen, J. Tian, and H. Xiong. 2012. An unsupervised approach to modeling personalized contexts of mobile users. *Knowledge and Information Systems* 31, 2, 345–370.

L. E. Baum, T. Petrie, G. Soules, and N. Weiss. 1970. A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains. *Annals of Mathematical Statistics* 41, 1, 164–171.

P. Belimpasakis, K. Roimela, and Y. You. 2009. Experience explorer: A life-logging platform based on mobile context collection. In *Proceedings of the 3rd International Conference on Next Generation Mobile Applications, Services and Technologies (NGMAST'09)*. IEEE, 77–82.

D. M. Blei and J. D. Lafferty. 2006. Dynamic topic models. In *Proceedings of the 23rd International Conference on Machine Learning*. ACM, 113–120.

D. M. Blei, A. Y. Ng, and M. I. Jordan. 2003. Latent Dirichlet allocation. *Journal of Machine Learning Research* 3, 993–1022.

H. Cao, T. Bao, Q. Yang, E. Chen, and J. Tian. 2010. An effective approach for mining mobile user habits. In *Proceedings of the 19th ACM International Conference on Information and Knowledge Management*. ACM, 1677–1680.

H. Cao, D. Jiang, J. Pei, E. Chen, and H. Li. 2009. Towards context-aware search by learning a very large variable length hidden Markov model from search logs. In *Proceedings of the 18th International Conference on World Wide Web*. ACM, 191–200.

S. Chu. 2008. Unstructured audio classification for environment recognition. In *Proceedings of the 23rd AAAI Conference on Artificial Intelligence*. 13–17.

G. A. Churchill. 1989. Stochastic models for heterogeneous dna sequences. *Bulletin of Mathematical Biology* 51, 1, 79–94.

N. Eagle and A. S. Pentland. 2009. Eigenbehaviors: Identifying structure in routine. *Behavioral Ecology and Sociobiology* 63, 7, 1057–1066.

A. J. Eronen, V. T. Peltonen, J. T. Tuomi, A. P. Klapuri, S. Fagerlund, T. Sorsa, G. Lorho, and J. Huopaniemi. 2006. Audio-based context recognition. *IEEE Transactions on Audio, Speech, and Language Processing,* 14, 1, 321–329.

D. R. Fredkin and J. A. Rice. 1992. Bayesian restoration of single-channel patch clamp recordings. *Biometrics* 427–448.

S. Goldwater and T. Griffiths. 2007. A fully Bayesian approach to unsupervised part-of-speech tagging. In *Annual Meeting: Association for Computational Linguistics.* Vol. 45. 744.

S. Guha, L. Yi, and D. Neuberg. 2008. Bayesian hidden Markov modeling of array CGH data. *Journal of the American Statistical Association* 103, 485–497.

J. D. Hamilton. 1989. A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica: Journal of the Econometric Society*, 357–384.

G. Heinrich. 2005. Parameter Estimation for Text Analysis. Retrieved from http://www.arbylon.net/publications/text-est.pdf.

J. Himberg, K. Korpiaho, H. Mannila, J. Tikanmaki, and H. T. Toivonen. 2001. Time series segmentation for context recognition in mobile devices. In *Proceedings of the IEEE International Conference on Data Mining (ICDM'01)*. IEEE, 203–210.

R. Huang, G. Yu, and Z. Wang. 2013. Dirichlet process mixture model for document clustering with feature partition. *IEEE Transactions on Knowledge and Data Engineering* 25, 8, 1748–1759.

H. Ishwaran and L. F. James. 2001. Gibbs sampling methods for stick-breaking priors. *Journal of the American Statistical Association* 96, 453, 161–173.

B.-H. Juang and L. R. Rabiner. 1991. Hidden Markov models for speech recognition. *Technometrics* 33, 3, 251–272.

P. Korpipää, M. Koskinen, J. Peltola, S.-M. Mäkelä, and T. Seppänen. 2003. Bayesian approach to sensor-based context awareness. *Personal and Ubiquitous Computing* 7, 2, 113–124.

C. Leggetter and P. Woodland. 1995. Maximum likelihood linear regression for speaker adaptation of continuous density hidden Markov models. *Computer Speech and Language* 9, 2, 171.

T. Lemlouma and N. Layaïda. 2004. Context-aware adaptation for mobile devices. In *Proceedings of the 2004 IEEE International Conference on Mobile Data Management*. IEEE, 106–111.

B. G. Leroux and M. L. Puterman. 1992. Maximum-penalized-likelihood estimation for independent and Markov-dependent mixture models. *Biometrics* 545–558.

X. Li, H. Cao, E. Chen, and J. Tian. 2012. Learning to infer the status of heavy-duty sensors for energy-efficient context-sensing. *ACM Transactions on Intelligent Systems and Technology (TIST)* 3, 2, 35.

N. Liampotis, N. Kalatzis, I. Roussaki, P. Kosmides, I. Papaioannou, E. Sykas, M. Anagnostou, and S. Xynogalas. 2012. Addressing the context-awareness requirements in personal smart spaces. In *Proceedings of the 2012 IEEE Asia-Pacific Services Computing Conference (APSCC'12)*. IEEE, 281–285.

L. Liao, D. J. Patterson, D. Fox, and H. Kautz. 2007. Building personal maps from gps data. *Annals of the New York Academy of Sciences* 1093, 1, 249–265.

J. S. Liu, A. F. Neuwald, and C. E. Lawrence. 1999. Markovian structures in biological sequence alignments. *Journal of the American Statistical Association* 94, 445, 1–15.

H. Ma, H. Cao, Q. Yang, E. Chen, and J. Tian. 2012. A habit mining approach for discovering similar mobile users. In *Proceedings of the 21st International Conference on World Wide Web*. ACM, 231–240.

R. M. Neal. 2000. Markov chain sampling methods for Dirichlet process mixture models. *Journal of Computational and Graphical Statistics* 9, 2, 249–265.

B. Nham, K. Siangliulue, and S. Yeung. 2008. Predicting mode of transport from iPhone accelerometer data. Technical report, Stanford University.

K. Nigam, A. K. McCallum, S. Thrun, and T. Mitchell. 2000. Text classification from labeled and unlabeled documents using em. *Machine Learning* 39, 2, 103–134.

B. Peng, Y. Wang, and J.-T. Sun. 2012. Mining mobile users activities based on search query text and context. In *Advances in Knowledge Discovery and Data Mining*. Springer, 109–120.

A. Pievatolo, F. Ruggeri, and R. Soyer. 2012. A Bayesian hidden Markov model for imperfect debugging. *Reliability Engineering & Systems Safety* 103, 11–21.

L. R. Rabiner. 1989. A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE* 77, 2, 257–286.

N. Ravi, N. Dandekar, P. Mysore, and M. L. Littman. 2005. Activity recognition from accelerometer data. In *Proceedings of the National Conference on Artificial Intelligence*. Vol. 20. AAAI Press, 1541.

R. Rawassizadeh, M. Tomitsch, K. Wac, and A. M. Tjoa. 2012. Ubiqlog: A generic mobile phone-based life-log framework. *Personal and Ubiquitous Computing*, 1–17.

P. Resnik and E. Hardisty. 2010. Gibbs sampling for the uninitiated. Technical report, DTIC Document.

B. Schilit, N. Adams, and R. Want. 1994. Context-aware computing applications. In *Proceedings of the 1st Workshop on Mobile Computing Systems and Applications (WMCSA'94).* IEEE, 85–90.

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

T. Teraoka. 2011. A study of exploration of heterogeneous personal data collected from mobile devices and web services. In *Proceedings of the 5th FTRA International Conference on Multimedia and Ubiquitous Engineering (MUE'11)*. IEEE, 239–245.

K. Yu, B. Zhang, H. Zhu, H. Cao, and J. Tian. 2012. Towards personalized context-aware recommendation by mining context logs through topic models. In *Proceedings of the 16th Pacific-Asia Conference on Advances in Knowledge Discovery and Data Mining*. Springer-Verlag, Kuala Lumpur, Malaysia, 431–443.

Y. Zheng, L. Liu, L. Wang, and X. Xie. 2008. Learning transportation mode from raw GPS data for geographic applications on the web. In *Proceedings of the 17th International Conference on World Wide Web*. ACM, 247–256.

H. Zhu, H. Cao, E. Chen, H. Xiong, and J. Tian. 2012a. Exploiting enriched contextual information for mobile app classification. In *Proceedings of the 21st ACM International Conference on Information and Knowledge Management*. ACM, 1617–1621.

H. Zhu, E. Chen, K. Yu, H. Cao, H. Xiong, and J. Tian. 2012b. Mining personal context-aware preferences for mobile users. In *Proceedings of the IEEE 12th International Conference on Data Mining (ICDM'12)*. IEEE, 1212–1217.

Received July 2013; revised December 2013; accepted April 2014