Mining Mobile User Preferences for Personalized Context-Aware Recommendation

HENGSHU ZHU, University of Science and Technology of China
ENHONG CHEN, University of Science and Technology of China
HUI XIONG, Rutgers University
KUIFEI YU, Nokia Research Center
HUANHUAN CAO, Nokia Research Center
JILEI TIAN, Nokia

Recent advances in mobile devices and their sensing capabilities have enabled the collection of rich contextual information and mobile device usage records through the device logs. These context-rich logs open a venue for mining the personal preferences of mobile users under varying contexts and thus enabling the development of personalized context-aware recommendation and other related services, such as mobile online advertising. In this paper, we illustrate how to extract personal context-aware preferences from the context-rich device logs, or context logs for short, and exploit these identified preferences for building personalized context-aware recommender systems. A critical challenge along this line is that the context log of each individual user may not contain sufficient data for mining his/her context-aware preferences. Therefore, we propose to first learn common context-aware preferences from the context logs of many users. Then, the preference of each user can be represented as a distribution of these common context-aware preferences. Specifically, we develop two approaches for mining common context-aware preferences based on two different assumptions, namely, context independent and context dependent assumptions, which can fit into different application scenarios. Finally, extensive experiments on a real-world data set show that both approaches are effective and outperform baselines with respect to mining personal context-aware preferences for 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-Aware, Personalized Recommendation, Mobile User

1. INTRODUCTION

With the rapid development of mobile industry, the mobile platform, such as smartphones and tablets, has become one of the most important media for social entertainment and information acquisition. Meanwhile, the advances in mobile devices enable them to equip with a rich set of context sensors, such as GPS sensors, 3D accelerometers, and optical sensors. These context sensors can capture the rich contextual information of mobile users, and thus enable a wide range of context-aware services, such as context-aware tour guide [Emrich et al. 2009], location based reminder [Sohn et al. 2009]. This is a substantially extended and revised version of [Zhu et al. 2012b], which appears in Proceedings of the 12th IEEE International Conference on Data Mining (ICDM’2012).

Author’s addresses: H. Zhu and E. Chen, School of Computer Science and Technology, University of Science and Technology of China; H. Xiong, Rutgers University; K. Yu and H. Cao and J. Tian, Beijing Lab of Nokia Research Center.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies show this notice on the first page or initial screen of a display along with the full citation. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, to republish, to post on servers, to redistribute to lists, or to use any component of this work in other works requires prior specific permission and/or a fee. Permissions may be requested from Publications Dept., ACM, Inc., 2 Penn Plaza, Suite 701, New York, NY 10121-0701 USA, fax +1 (212) 869-0481, or permissions@acm.org.

© YYYY ACM 2157-6904/YYYY/01-ARTA $15.00
DOI: http://dx.doi.org/10.1145/000000.000000

ACM Transactions on Intelligent Systems and Technology, Vol. V, No. N, Article A, Publication date: January YYYY.
2005], and context-aware recommendation [Bader et al. 2011; Liu et al. 2013; jae Kim et al. 2010; Karatzoglou et al. 2010; Woerndl et al. 2007]. In fact, the contextual information and corresponding usage records (e.g., browsing web sites and playing games) can be recorded in context-rich device logs, or context logs for short, which can be used for mining the personal context-aware preferences of mobile users, i.e., which category of contents is preferred by a particular user under a certain context. Particularly, mining such preferences is a fundamental work for understanding the behaviors of mobile users, and thus enables better context-aware services. Specifically, by considering both the personal context-aware preferences and the current contexts of users, it is possible to provide personalized context-aware recommendation and other related services for mobile users, such as mobile online advertising. Figure 1 shows some examples of how to exploit contextual information for recommendation services. Indeed, the personalized context-aware recommendation services can provide better user experiences than traditional context-aware recommender systems which only consider the contextual information but not different users’ preferences under the same context [Zhu et al. 2012b]. For instance, the following two examples intuitively illustrate how the context-aware recommendation services can improve the user experiences.

**Example 1.1 (Context-Aware Content Pushing).** Content pushing is an important function of smart mobile operating systems, which aims to automatically deliver the right content to mobile users and is widely used for content recommendation and intelligent advertising [Podnar et al. 2002]. Particularly, suppose that a mobile user Kate would like to play her mobile phone while taking the bus. Through the analysis of the sensing information collected on her smart phone, a context-aware recommender system could discover that Kate is taking a bus (sensed by 3D accelerometers [Kwapisz et al. 2011]) from the work place to home (sensed by GPS or cell ID combined with data mining on historical trajectories of users [Eagle et al. 2009; Zheng et al. 2009]) in a Monday evening (sensed by the system clock). Therefore, with respect to Kate’s personal context-aware preference, e.g., Kate often listens to R&B music under the same context, the recommender system may push some new R&B music and related advertisements to Kate.

**Example 1.2 (Context-Aware Mobile Advertising).** Mobile advertising, such as banner or in-App advertising, has become one of the most important components of online advertising campaigns. In fact, a crucial need of mobile advertisers is to understand the behavior and preferences of mobile users. Therefore, with the help of the personalized context-aware recommender system, the advertiser can better design and deliver advertisements to mobile users. For example, through mining users context logs,
the recommender system may discover that a mobile user Tom has two typical context-aware preferences “playing games every evening at home” and “using restaurant review Apps every workday afternoon near office”, thus it can suggest advertisers to deliver advertisements of popular games and restaurants to Tom when he is using mobile phone in corresponding contexts. Furthermore, such context-aware preferences can also be used for re-ranking the sponsored advertisements with respect to users’ different contexts.

In recent years, although many researchers studied the problem of personalized context-aware recommendation [Zheng et al. 2010; Park et al. 2007; Ge et al. 2011; Liu et al. 2011; jae Kim et al. 2010] and proposed some approaches for mining personal context-aware preferences, most of them did not take into account context-rich information in their approaches. Also, some of these studies are based on item ratings generated by users under different contexts, which are difficult to obtain in practice. In contrast, usage records in context-rich device logs are a rich resource for mining personalized context-aware user preferences. However, it is still under-addressed about how to mine context-aware preferences from context-rich logs for developing context-aware recommender systems.

To this end, in this paper, we propose a novel approach for mining personal context-aware preferences from context-rich device logs of mobile users. A critical challenge for mining personal context-aware preferences is that the context log of each individual user usually does not contain sufficient training information. As a result, it can be difficult to learn personal context-aware preferences if we only use the context log of each individual user. Therefore, we propose to first find common context-aware preferences from the context logs of many users and then represent the context-aware preference of each user as a distribution of common context-aware preferences. Moreover, on the basis of two different assumptions about context data dependency, we propose two methods for mining common context-aware preferences. The first one is more efficient but sacrifices a little performance, while the second one needs more training time but has better performances. Figure 2 illustrates the overview of the proposed approach and how the mined preferences are used for predicting the preferred categories of contents for a given mobile user under a certain context. Specifically, the contributions of this paper are summarized as follows.

ACM Transactions on Intelligent Systems and Technology, Vol. V, No. N, Article A, Publication date: January YYYY.
First, we propose a novel approach for mining the personal context-aware preferences for mobile users through the analysis of context-rich device logs. Specifically, we propose to first mine common context-aware preferences from the context logs of many users and then represent the personal context-aware preference of each user as a distribution of common context-aware preferences. The mined personal context-aware preferences can enable the development of personalized context-aware recommender systems and other related services, such as mobile online advertising.

Second, we design three effective methods for mining common context-aware preferences based on two different assumptions about context data dependency. If context data are assumed to be conditionally independent, we propose to mine common context-aware preferences through topic models. Otherwise, if context data are assumed to be dependent, we propose to exploit the constraint based Matrix Factorization techniques for mining common context-aware preferences and only consider those contexts which are relevant to content usage for reducing the computation complexity.

Finally, we evaluate the proposed approach using a real-world data set with context logs collected from 443 mobile phone users. In total, there are more than 8.8 million context records. This data set contains much more context-rich information and is much bigger than those reported in previous works on context log mining [Bao et al. 2010; Cao et al. 2010]. The experimental results clearly demonstrate the effectiveness of the proposed approach and indicate some inspiring findings.

Overview. The remainder of this paper is organized as follows. In Section 2, we introduce the details about context logs and give the overview of the proposed approach. Section 3 and Section 4 present our novel approaches for mining common context-aware preferences of many users on the basis of two different assumptions, respectively. In Section 5, we report the experimental results on a real-world data set. Section 6 provides a brief review of related works. Finally, in Section 7, we conclude the paper.

2. MINING PERSONAL CONTEXT-AWARE PREFERENCES FROM CONTEXT LOGS

Smart devices can capture the historical context data and the corresponding usage records of users through multiple sensors and record them in context logs. For example, Table I shows a toy context log, which contains several context records, and each context record consists of a timestamp, the most detailed context at that time, and the corresponding usage record. A context consists of several contextual features (e.g., Day name, Time range, and Location) and their corresponding values (e.g., Saturday, AM8:00-9:00, and Home), which can be annotated as contextual feature-value pairs. Moreover, usage records can be empty (denoted as “Null”) because a user does not always use the mobile phone.

Note that, in Table I, raw locations in context data, such as GPS coordinates or cell IDs, have been transformed into semantic locations such as “Home” and “Work Place” by some location mining approaches (e.g., [Eagle et al. 2009]). The basic idea of these approaches is to find the clusters of user locations and recognize their semantic meaning by a time pattern analysis. Moreover, we also map the raw usage records to the usage records of particular categories of contents via some mobile App classification approaches (e.g., [Zhu et al. 2012a]). For example, we can map two raw usage records “Play Angry Birds” and “Play Fruit Ninja” to the usage records of content category “Action Games”.

In this way, the context data and usage records in context logs are normalized and the data sparseness problem is somewhat alleviated. The above helps the task of personal context-aware preference mining.

Intuitively, context logs contain rich information about content usage given particular contexts and can be used for mining the personal context-aware preferences of
null
but can simplify the problem. For example, under such an assumption, given a context “{(Time range: PM10:00-11:00), (Location: Home)}” and a mobile user $u$, if we can infer the latent common context-aware preference distribution of $u$, we only need to consider which category of contents $u$ may prefer under the context (Time range: PM10:00-11:00) and the context (Location: Home) given each common context-aware preference, but not need to consider which category of contents $u$ may prefer given the co-occurrence of (Time range: PM10:00-11:00) and (Location: Home) given each common context-aware preference.

The second assumption is that different types of context data are mutually dependent, which is relatively weak and may be more proper in practice. However, such an assumption makes it more difficult for modeling context-aware preferences. For example, under such an assumption, given the above context, we have to consider the co-occurrence of (Time range: PM10:00-11:00) and (Location: Home) when making a preference prediction. Obviously, the corresponding models may be more complex than the ones based on the first assumption. In this paper, we propose different approaches based on the above two assumptions and conduct extensive experiments to evaluate them. The details of our novel approaches based on two different assumptions are presented in the following two sections, respectively.

### 3. CONTEXT-AWARE PREFERENCE MINING BASED ON CONTEXT CONDITIONAL INDEPENDENCY ASSUMPTION

We first propose a method based on the assumption that different types of context data are conditional independent given a particular common context-aware preference. Under such an assumption, given a context $C = \{p_1, p_2, ..., p_l\}$ where $p_i$ denotes an atomic context, i.e., a contextual feature-value pair, the probability that a user $u$ prefers content category $c$ can be represented as

$$P(c|C, u) \propto \sum_z \left( P(c, C|z) \cdot P(z|u) \right) \propto \sum_z \left( \prod_{p_i \in C} P(c, p_i|z) \cdot P(z|u) \right)$$

Therefore, the problem is further converted to learn $P(c, p_i|z)$ and $P(z|u)$ from many users’ context logs, which can be solved by widely used topic models. In this section, we present how to utilize topic models for mining common context-aware preferences by estimating $P(c, p_i|z)$ and $P(z|u)$. For simplicity, we refer to the co-occurrence of a usage of a content in category $c$ and the corresponding contextual feature-value pair $p_i$, i.e., $(c, p_i)$, as an Atomic Context-aware Preference feature, and ACP-feature
3.1. Mining Common Context-Aware Preferences through Topic Models

Topic models are generative models that are successfully used for document modeling. They assume that there exist several topics for a corpus $D$ and a document $d_i$ in $D$ can be taken as a bag of words $\{w_{i,j}\}$ which are generated by these topics. Intuitively, if we take ACP-features as words, take context logs as bags of ACP-features to correspond documents, and take common context-aware preferences as topics, we can take advantage of topic models to learn common context-aware preferences from many users’ context logs.

Since raw context logs are not naturally in the form of bags of ACP-features, we need to extract bags of ACP-features from them as training data. Specially, we first remove all context records without any usage record and then extract ACP-feature from the remaining ones. Given a context record $< Tid, C, c >$ where $Tid$ denotes a timestamp, $C = \{p_1, p_2, ..., p_l\}$ denotes a context and $c$ denotes the category of the used content in the usage record, we can extract $l$ ACP-features, namely, $(c, p_1), (c, p_2), ..., (c, p_l)$. For simplicity, we refer the bag of ACP-features extracted from user $u$’s context log as the ACP-feature bag of $u$. Figure 3 shows the example of generation process of ACP-feature bag from context records.

3.1.1. LDA based Context Modeling. Among several existing topic models, in this paper, we first leverage the widely used Latent Dirichlet Allocation model (LDA) [Blei et al. 2003]. According to LDA, the ACP-feature bag of user $u_i$ denoted as $d_i$ is generated as follows. First, before generating any ACP-feature bag, $K$ prior ACP-feature conditional distributions given context-aware preferences $\{\phi_z\}$ are generated from a prior Dirichlet distribution $\beta$. Secondly, a prior common context-aware preference distribution $\theta_i$ is generated from a prior Dirichlet distribution $\alpha$ for each user $u_i$. Then, for generating the $j$-th ACP-feature in $d_i$ denoted as $w_{i,j}$, the model firstly generates a common context-aware preference $z$ from $\theta_i$ and then generates $w_{i,j}$ from $\phi_z$. Figure 4 (a) shows the graphic representation of modeling ACP-feature bags by LDA model.

The process of LDA model training is to learn the proper latent variables $\theta$ and $\phi$ to maximize the posterior distribution of the observed ACP-feature bags, i.e., $P(u|\alpha, \beta, \theta, \phi)$. In this paper, we take advantage of a Markov chain Monte Carlo method named Gibbs sampling [Griffiths and Steyvers 2004] for training LDA models. This method begins with a random assignment of common context-aware preferences to ACP-features for initializing the state of Markov chain. In each of the following iterations, the method will re-estimate the conditional probability of assigning a common context-aware preference to each ACP-feature, which is conditional on the assignment. 

Fig. 4. The graphic representation of modeling ACP-feature bags by (a) LDA topic model, and (b) LDAC topic model.

for short. For example, $(IsHoliday? : Yes) \rightarrow Games$ is an ACP-Feature, where $p = (IsHoliday? : Yes)$ and $c = Games$. 

ACM Transactions on Intelligent Systems and Technology, Vol. V, No. N, Article A, Publication date: January YYYY.
of all other ACP-features. Then a new assignment of common context-aware preferences to ACP-features according to those latest calculated conditional probabilities will be scored as a new state of Markov chain. Finally, after rounds of iterations, the assignment will converge, which means each ACP-feature is assigned a stable and final common context-aware preference and we can obtain the estimation of \( P(z|u) \) and \( P(c, p|z) \) as follows.

\[
P(z|u) = P(z|d_u) = \frac{n_{u,z} + \alpha_z}{\sum_i^K n_{u,z_i} + \sum_i^K \alpha_{z_i}},
\]

\[
P(c, p|z) = \frac{n_{z,c,p} + \beta_{c,p}}{\sum_i^M n_{z,(c,p)_i} + \sum_i^M \beta_{(c,p)_i}},
\]

where the \( n_{u,z} \) is the frequency of ACP-features in \( u \) have been assigned to common context-aware preference \( z \), and \( n_{z,(c,p)_i} \) is the frequency of \( i \)-th ACP-feature has been assigned to common context-aware preference \( z \).

3.1.2. LDAC based Context Modeling. Although LDA can model the ACP-feature bags in an intuitive way, from some real-world observations we find the generation of ACP-features are not only decided by latent common context-aware preferences but also by their internal contextual features. For example, BlueTooth information can only be got when user opens BlueTooth sensor, and location information often cannot be obtained in underground subways due to the lack of GPS/cell ID information. Therefore, to more accurately model context information, in this paper we also leverage the extended Latent Dirichlet Allocation on Context Model (LDAC) [Bao et al. 2010] for mining latent common context-aware preferences.

In the LDAC model, the ACP-feature bags of user \( u_i \) denoted as \( d_i \) is generated as follows. Firstly, a prior common context-aware preference distribution \( \pi_{d_i} \) is generated from a prior Dirichlet distribution \( \alpha \). Secondly, a prior contextual feature distribution \( \pi_{d_i,j} \) is generated from a prior Dirichlet distribution \( \gamma \). Then, for the \( j \)-th ACP-feature in \( d_i \), a common context-aware preference \( z_{d_i,j} \) is generated from \( \pi_{d_i,j} \), a contextual feature \( f_{d_i,j} \) is generated from \( \pi_{d_i,j} \), and the content category with contextual feature value of \( f_{d_i,j} \) denoted as \( (c, v_{d_i,j}) \) is generated from the distribution \( \phi_{z_{d_i,j}, f_{d_i,j}} \). Moreover, there are totally \( K \times F \) prior distributions of ACP-features \( \{ \phi_k \} \) which follow a Dirichlet distribution \( \beta \), where \( F \) is the number of unique contextual features. Figure 4 (b) shows the graphical representation of the LDAC model. The process of LDAC model training is to learn the proper latent variables to maximize the posterior distribution of the observed ACP-feature bags, i.e., \( P(d, \theta, z_d, \pi_d, \Phi|\alpha, \beta, \gamma) \).

In this paper, we leverage the Gibbs sampling based approach introduced in [Bao et al. 2010] to train LDAC model. After training process, we can obtain the probabilities \( P(z|u) \) by Equation 1 and \( P(c, p|z) \) as follows \(^1\).

\[
P(c, p|z_k) = P(c, v_p|f_p, z)P(f_p),
\]

where \( p = (f_p : v_p) \) and

\[
P(c, v_p|f_p, z) = \frac{n_{z,c,f_p,v_p} + \beta_{c,v_p}}{\sum_v n_{z,c,f_p,v} + \sum_{v \in f_m} \beta_v},
\]

\[
P(f_p) = \frac{\sum_z \sum_v n_{k,c,f_p,v} + \gamma_{f_m}}{\sum_f \sum_z \sum_v n_{k,c,f_p,v} + \sum_f \gamma_f},
\]

\(^1\)The detailed inference can be found in Appendix.
where \( n_{z,c,f_p,v_p} \) is the frequency of ACP-feature \((c,p)\) has been assigned to common context-aware preference \(z\), and \( V_{f_m} \) is the number of values for contextual feature \( f_m \).

3.2. Selecting the Number of Common Context-Aware Preferences

Both LDA and LDAC models need a predefined parameter \(K\) to determine the number of common context-aware preferences. In this paper we utilize the method proposed in [Bao et al. 2010] to estimate \(K\). To be specific, we first empirically define a topic number range \([K_{min}, K_{max}]\) and then select two groups of ACP-feature bags as the training set \(S_1\) and the test set \(S_2\), respectively. Then we can determine \(K\) by the corresponding perplexity [Azzopardi et al. 2003; Blei et al. 2003] of the test set \(S_2\). Here we take LDA model as example, the perplexity is defined as follows.

\[
Perplexity(S_2) = \exp\left\{ \frac{-\sum_{d_i \in S_2} \log \{ P(d_i|S_1) \}}{\sum_{d_i \in S_2} N_{d_i}} \right\},
\]

where \(N_{d_i}\) indicates the number of ACP-features in \(d_i\) and

\[
P(d_i|S_1) = \prod_{(c,p) \in d_i} P(c,p|S_1) = \prod_{(c,p) \in d_i} \sum_{j=1}^{K} \left( P(c,p|z_j) P(z_j|S_1) \right).
\]

Herein, \(P(c,p|z_j)\) can be obtained after model training, and \(P(z_j|S_1)\) can be estimated by \(n_{d_i,j}^a, n_{d_i,k}^a\), where \(n_{d_i,j}\) indicates the number of ACP-features labeled with \(z_j\) in \(d_i\).

The perplexity is widely used to index the modeling performance. The smaller the perplexity, the better modeling performance it implies. However, the perplexity may consistently drop with the increase of \(K\) in practical data sets. To avoid the over-fitting problem, we cannot only utilize the minimum perplexity as the metric for determining \(K\) [Azzopardi et al. 2003; Blei et al. 2003]. A complementary method is to define a decline rate \(\zeta\) of perplexity and stop seeking better \(K\) if the decline rate of perplexity is less than \(\zeta\). In our experiments, we set \(\zeta\) to be 10% according to [Bao et al. 2010].

4. CONTEXT-AWARE PREFERENCE MINING BASED ON CONTEXT DEPENDENCY ASSUMPTION

Since it may be relatively strong to assume that different types of context data are conditionally independent, we also propose a method for mining common context-aware preferences based on the assumption that different types of context data are mutually dependent. Under such an assumption, we cannot decompose contexts into atomic contexts and need to learn \(P(c,C|z)\) directly from user context logs. A major challenge is that we cannot learn all conditional distributions \(P(c,C|z)\) for all \(C\) simply because the number of unique \(C\) is exponential to the number of unique contextual feature-value pairs and we will suffer the assemble explosion problem if we learn all of them. Fortunately, we observe that usually not all parts of a context are relevant to content usage and thus the corresponding preferences of content categories. For example, given a context “\((\text{Day name: Sunday}), (\text{Day period: Evening}), (\text{Location: Home}), (\text{Battery level: Low})\)”, we may not be able to demonstrate the whole context is relevant to a particular content category through the analysis of context logs. Instead, we may find context logs show that some parts of them such as “\((\text{Day name: Sunday}), (\text{Day period: Evening})\)” and “\((\text{Day period: Evening}), (\text{Location: Home})\)” are indeed relevant to a particular content category. To this end, an intuitive idea is to only consider the content-relevant parts of contexts for predicting personalized context-aware preferences of content categories. These content-relevant parts can be referred to as content-relevant contexts for simplicity. Along this line, we only need to learn the conditional distributions...
$P(c, C^r|z)$ and $P(z|u)$, where $C^r$ denotes a content-relevant context. Moreover, given a context $C$, it can be divided into two cases to calculate $P(c, C|z)$.

First, if $C$ contains some content-relevant contexts, we can calculate $P(c, C|z)$ directly by its maximal sub-contexts which are also content-relevant contexts as follows.

$$P(c, C|z) = \frac{1}{|C^r_{\text{max}}|} \times \sum_{C^r_{\text{max}}} P(c, C^r_{\text{max}}|z),$$

where $C^r_{\text{max}}$ denotes a maximal content-relevant sub-context contained by $C$, and $|C^r_{\text{max}}|$ indicates the number of $C^r_{\text{max}}$. Take the above context $C$ = (Day name: Sunday), (Day period: Evening), (Location: Home), (Battery level: Low) for example, suppose that there are two maximum content-relevant sub-contexts $C^r_1$ = (Day name: Sunday), (Day period: Evening) and $C^r_2$ = (Day period: Evening), (Location: Home) in it, we will recommend corresponding contents to the given user by considering his/her preference under both $C^r_1$ and $C^r_2$ equally.

Second, if $C$ does not contain any content-relevant context, we can estimate $P(c, C|z)$ by normalizing the probabilistic space of the joint distribution of $c$ and $C$ conditional on $z$. Especially, we let

$$P(c, C^0|z) = \frac{1}{N_C \cdot N_c} \times \left(1 - \sum_{c} \sum_{C^\Delta} P(c, C^\Delta|z) \right),$$

where $C^0$ denotes a context without any content-relevant sub-context, $N_C$ indicates the total number of $C^0$, $N_c$ indicates the total number of unique content categories, and $C^\Delta$ denotes a context with at least one content-relevant sub-context. Actually, we do not need to calculate $P(c, C|z)$ in this case because it is the same with varying $c$ and cannot help to make a recommendation decision. In practice, we do not recommend any content in this case.

Therefore, the original problem is divided into two sub-problems, namely, how to discover those content-relevant contexts? and how to learn common context-aware preferences and user personal distributions of common context-aware preferences, i.e., $P(c, C^r|z)$ and $P(z|u)$? The solutions for the two sub-problems are presented in the following sections in detail, respectively.

4.1. Discovering the Content-Relevant Context

An intuitive way of discovering the context relevant to some content categories is mining association rules [Agrawal and Srikant 1994] between them with predefined minimum supports and minimum confidences. Therefore, given a content-relevant context $C^r$ and a content category $c$, $P(c|C^r, u) \propto P(c, C^r|u)$ can be calculated as $P(c, C^r|u) = P(c|C^r, u)P(C^r|u)$, where $P(c|C^r, u)$ can be estimated by the corresponding confidence of the association “$C^r \rightarrow c$” and $P(C^r|u)$ can be estimated by $\frac{\text{Support}(C^r)}{N_r}$, where $N_r$ indicates the total number of context records in the context log of user $u$.

However, as pointed out by Cao et al. [Cao et al. 2010], the amounts of context data and user activity records are usually extremely unbalanced, which makes it difficult to mine such association rules through traditional association rule mining approaches. An alternative approach is only leveraging the context records with non-empty activity records. However, it will lose the discriminative information on how likely no activity will be done under a particular context. Fortunately, some researchers have studied this problem and proposed some novel algorithms for mining such association rules. For example, Cao et al. [Cao et al. 2010] proposed a novel algorithm called GCPM (Generating Candidates for behavior Pattern Mining) for mining such association rules, which are referred to as behavior patterns in their work, by utilizing different ways
of calculating supports and confidences. In an incremental work of [Cao et al. 2010], Li et al [Li et al. 2012] proposed a more efficient algorithm named BP-Growth for this problem. In this paper, we leverage the BP-Growth algorithm for mining such association rules. The basic idea of the algorithm is partitioning the original context logs into smaller sub-context logs for reducing the mining space and mining frequent association rules in these sub-context logs. Table II illustrates some examples of association rules mined from context logs. In fact, these association rules clearly demonstrate the personal context-aware preferences of mobile users. For example, the context “(Time Period: Evening) (Location: Home)” may imply a leisure time at home, thus the user has a preference “Web” under such context.

It is worth noting that the mining is performed on individual users’ context logs because merging all context logs may normalize the associations between contexts and content categories. For example, given several users who usually play action games in bus and several users usually play other games in bus. If we try to mine the associations between contexts and content category by merging all users’ context logs, we may falsely conclude that “In bus” has no significant relevance with any content category. In contrast, we can discover “In bus” is both relevant to action games and other games according to different people by taking into account each user’s context log separately.

4.2. Mining Common Context-Aware Preferences through Constraint based Bayesian Matrix Factorization

After finding content-relevant contexts, the remaining task is to learn common context-aware preferences and user personal distributions of common context-aware preferences, i.e., $P(c, C'|z)$ and $P(z|u)$. By building a matrix of $P(c, C'|u)$, where each column denotes a probabilistic distribution of different $(c, C')$ pairs for a given user $u$, we can convert this task into a matrix factorization problem as follows.

$$\Omega_{N \times M} = \Phi_{N \times K} \times \Theta_{K \times M} + N_{N \times M},$$

where $N$ indicates the number of unique $(c, C')$ pairs, $M$ indicates the number of users and $K$ indicates the number of common context-aware preferences. To be specific, $\Omega$ denotes the observed matrix of $P(c, C'|u)$, $\phi_{ik} \in \Phi(1 \leq i \leq N, 1 \leq k \leq K)$ denotes the probability $P(c, C'|z_k)$, $\theta_{kj} \in \Theta(1 \leq k \leq K, 1 \leq j \leq M)$ denotes the probability $P(z_k|u_j)$, and the matrix $N$ denotes the residual noise information. Moreover, the matrix factorization task has two additional constraints for possible solutions as follows: 1) all elements in matrix $\Phi$ and $\Theta$ should be non-negative values, 2) $\forall j: 1 \leq j \leq M \sum_{k=1}^{K} \theta_{k,j} = 1$ and $\forall k: 1 \leq k \leq K \sum_{i=1}^{N} \phi_{i,k} = 1$, which are both obvious since each column of $\Phi$ and $\Theta$ denotes a probabilistic distribution. Figure 5 shows the process of mining common context-aware preferences through constrained matrix factorization.

According to the problem statement and constraints above, the objective of our matrix factorization task is to find a possible solution for matrix $\Phi$, $\Theta$ and $N$. In this paper, we propose to leverage a constraint based Bayesian Matrix Factorization model [Schmidt 2009] for resolving this problem. In this model, we can perform matrix factorization with multiple inequality and equality constraints. Specifically, we aim...
to infer the posterior probabilistic distributions of $\Phi$ and $\Theta$ under a set of model assumptions, which are specified by the likelihood function $P(\Omega | \Phi, \Theta, N)$. The likelihood function denotes the probability of the observed data matrix $\Omega$ given priors $P(\Phi, \Theta)$ and $P(N)$. According to [Schmidt 2009], to perform efficient inference based on Gibbs sampling, we select priors as follows. First, we select an i.i.d. zero mean Gaussian noise model as follows.

$$P(n_{ij}) = N(n_{ij} | 0, \nu_{ij}) = \frac{1}{\sqrt{2\pi\nu_{ij}}} \exp\left(-\frac{n_{ij}^2}{2\nu_{ij}}\right),$$

where parameter $\nu_{ij}$ satisfies conjugate inverse-gamma prior that satisfies,

$$P(\nu_{ij}) = IG(\nu_{ij} | \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} \nu_{ij}^{-(\alpha+1)} \exp\left(-\frac{\beta}{\nu_{ij}}\right).$$

Then, we select a Gaussian prior over $\Phi$ and $\Theta$ subject to inequality constraints $Q$ and equality constraints $R$ as

$$P(\phi, \theta) \propto \begin{cases} \begin{bmatrix} [\phi] & [\theta] \\ \Sigma_{\phi} & \Sigma_{\theta} \\ \Sigma_{\phi}^T & \Sigma_{\theta}^T \\ \mu_{\phi} & \mu_{\theta} \end{bmatrix} & \text{if } Q(\phi, \theta) \leq 0, \\
0 & \text{otherwise,} \end{cases}$$

where $\phi = (\phi_{11}, \phi_{12}, ..., \phi_{NK})^T$ and $\theta = (\theta_{11}, \theta_{12}, ..., \theta_{KM})^T$.

With above definitions, we can utilize Gibbs sampling methods to estimate the posterior distributions as follows. In the first round of sampling, we randomly assign values for $\phi$ and $\theta$ according to the two constraints to initialize the state of Markov chain. Then, we calculate the density of noise variance $P(\nu_{ij} | \phi, \theta)$ by inverse-gamma distribution due to the choice of conjugate prior (i.e., Equation 4). Next, we can estimate $P(\phi | \Omega, \theta, N)$ and $P(\theta | \Omega, \phi, N)$ from the constraint Gaussian density (i.e., Equation 5). Finally, we re-generate values for $\phi$ and $\theta$ according to the new posterior probabilities to score a new state of Markov chain. After many rounds of iterations, the results of matrixes $\Phi$ and $\Theta$ will converge.

### 4.3. Selecting the Number of Common Context-Aware Preferences

As mentioned in Section 3, an important problem for mining common context-aware preferences is to select a proper number of common context-aware preferences. For
the Matrix Factorization based approach for mining common context-aware preferences, we utilize the Chib’s method introduced in [Schmidt et al. 2009] to infer the proper number of common context-aware preferences. To be specific, in the Bayesian framework, model selection can be performed by evaluating the marginal likelihood \( P(\Omega) \). The Chib’s method is based on the Bayes relation \( P(\Omega) = \frac{P(\Theta|\Omega) P(\Omega)}{P(\Theta)} \), where the numerator can be easily estimated by the trained model and the key problem is to estimate \( P(\Theta) \). Denoting each row of \( \Theta \) as \( \Theta_i \), we can calculate the denominator as \( P(\Theta|\Omega) = P(\Theta_1|\Omega) \times P(\Theta_2|\Theta_1, \Omega) \times \cdots \times P(\Theta_K|\Theta_1, \cdots, \Theta_{K-1}, \Omega) \). After \( R \) rounds of Gibbs sampling in our approach, we can estimate each term by averaging over the conditional density \( P(\Theta_K|\Theta_1, \cdots, \Theta_{K-1}, \Omega) \approx \frac{1}{R} \sum_{r=1}^{R} P(\Theta_K|\Theta_1, \cdots, \Theta_{K-1}, \Theta^{(r)}_1, \cdots, \Theta^{(r)}_{K-1}, \Omega) \), where \( \Theta^{(r)}_1, \cdots, \Theta^{(r)}_K \) are Gibbs samples from \( P(\Theta_1, \cdots, \Theta_K|\Theta_1, \cdots, \Theta_{K-1}, \Omega) \). Thus, given a range \([K_{min}, K_{max}]\) for the number of common context-aware preferences, we can select the \( K \in [K_{min}, K_{max}] \) which maximizes the likelihood.

5. EXPERIMENTAL RESULTS

In this section, we evaluate the performances of the three implementations of the proposed framework based on two kinds of contextual assumptions for predicting user preferences of content categories, namely, CIAP-LDA (Context conditional Independence Assumption based Prediction with LDA model), CIAP-LDAC (Context conditional Independence Assumption based Prediction with LDAC model) and CDAP (Context Dependency Assumption based Prediction), with several baseline methods on a real-world data set.

Table III. The types of contextual information in our data set.

<table>
<thead>
<tr>
<th>Context</th>
<th>Value range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Week</td>
<td>{Monday, Tuesday, ..., Sunday}</td>
</tr>
<tr>
<td>Is a holiday?</td>
<td>{Yes, No}</td>
</tr>
<tr>
<td>Day period</td>
<td>{Morning(7:00-11:00), Noon(11:00-14:00), Afternoon(14:00-18:00), Evening(18:00-21:00), Night(21:00-Next day 7:00)}</td>
</tr>
<tr>
<td>Time range</td>
<td>{0:00-1:00, 1:00-2:00, ..., 23:00-24:00}</td>
</tr>
<tr>
<td>Profile type</td>
<td>{General, Silent, Meeting, Outdoor, Pager, Offline}</td>
</tr>
<tr>
<td>Battery Level</td>
<td>{Level 1, Level 2, ..., Level 7}</td>
</tr>
<tr>
<td>Charging State</td>
<td>{Charging, Complete, Not Connected}</td>
</tr>
<tr>
<td>Social location</td>
<td>{Home, Work Place, On the way}</td>
</tr>
</tbody>
</table>

Fig. 6. (a) The distribution of 618 unique contents w.r.t. the corresponding content categories and (b) the distribution of context records w.r.t. the content categories their corresponding usage records belong to.

ACM Transactions on Intelligent Systems and Technology, Vol. V, No. N, Article A, Publication date: January YYYY.
5.1. Experimental Data

The data set used in the experiments is collected from many volunteers by a major manufacturer of smart mobile devices. The data set consists of 8,852,187 context records which contain rich contextual information and usage records of 443 smart phone users spanning from several weeks to several months. Table III shows the concrete types of context data the data set contains. In the experiments, we classified the 665 unique contents appearing in raw usage records into 12 content categories based on the taxonomy of Nokia Ovi store (www.ovi.com), which are Call, Web, Multimedia, Management, Game, System, Navigation, Business, Reference, Social Network Service (SNS), Utility and Others. In our experiments, we do not utilize the categories Call and others because they are not useful for generating corresponding recommendations. Instead, we only utilize other 10 content categories which contain 618 unique contents appearing in total 408,299 usage records.

Figure 6 (a) and (b) show the distribution of 618 unique contents in raw usage records with respect to the corresponding content categories and the distribution of context records with respect to the content categories their corresponding usage records belong to, respectively. The context records with empty usage records are not taken into account.

Figure 7 (a) compares the distributions of all context records and the context records with non-empty usage records for all users. From the figure we can see that usually though many context records of individual mobile users are collected, only a small proportion of them have non-empty usage records and can be used as training data, which implies the limit of mining personal context-aware preferences only from individual users’ context logs. Moreover, Figure 7 (b) shows the coverage ratio of unique contexts in each user’s context log compared with all unique contexts, from which we can see that the unique contexts in each individual user’s context log is relatively limited, which motivates learning from many users’ context logs as well.

5.2. Benchmark Methods

First, we select two baseline methods which are transformed from our approaches but only consider individual users’ context logs for evaluating whether only considering individual users’ context logs is not enough for the personalized context-aware preference mining.

CIAP-i stands for a variant of CIAP which only leverages individual users’ context logs. To be specific, in this approach, given a user $u$ and a context $C$, we predict the user preferred content categories by ranking each content category $c$ according to $P(c|u, C)$,
which can be estimated by $P(c|u, C) \propto \prod_{p \in C} P(c, p|u)$, and $P(c, p|u)$ can be calculated by $P(c, p|u) = \frac{n_{c,p}}{n_{c}}$, where $n_{c,p}$ and $n_{c}$ indicate the numbers of ACP-feature $(c, p)$ and all ACP-features appeared in the context log of $u$, respectively.

**CDAP-i** stands for a variant of CDAP which only leverages individual users’ context logs, in which given a user $u$ and a context $C$, the user preferred content categories are predicted by ranking each content category $c$ according to $P(c|u, C)$. If $C$ contains some personal content-relevant contexts denoted $C^r$, where $C^r$ should appear in some behavior patterns of $u$, we let $P(c|u, C) \propto P(c, C|u) \times \frac{1}{|C^r_{\text{max}}|} \sum_{C^r_{\text{max}}} P(c|C^r_{\text{max}}, u)$, where $C^r_{\text{max}}$ denotes a maximal personal content-relevant sub-context contained by $C$, $P(c|C^r_{\text{max}}, u)$ can be estimated by the confidence of the association “$C^r_{\text{max}} \rightarrow a$” in the context log of user $u$. Otherwise, if $C$ does not contain any personal content-relevant sub-context, we do not recommend any content to user $u$.

Then, we select a state-of-the-arts personalized context-aware recommendation approach based on individual users’ context logs as a baseline.

**CASVM** stands for personalized Context-Aware preference prediction by Ranking SVM, which is introduced in [Kahng et al. 2011]. To be specific, in this approach, given a user's $u$ and a context $C$, we calculate five types of features $P(c)$, $P(c|u)$, $P(c|p)$, $P(c|u, p)$ and $P(c|u, p_1, ..., p_n)$, where $p$ denotes a contextual feature-value pair in $C$, according to the context log of $u$ and then leverage Ranking SVM for ranking content category $c$.

Moreover, to validate the performance of leveraging many users’ context logs for mining personal context-aware preferences, we also select two state-of-the-arts collaborative filtering (CF) based approaches as baselines.

**CACF** stands for Context-Aware preference mining by disjunction CF, which is a memory-based CF approach introduced in [Lee et al. 2010]. In this approach, the preference score $s_{c,u,C}$ of the user $u$ on content category $c$ in the context $C$ is calculated by the disjunctive aggregation of the estimated preferences of $s_{c,u,p}$, where $p$ is a feature-value pair in $C$ and $s_{c,u,p}$ is measured by counting how many context records of $u$ contain $p$ and contents in category $c$.

**CATF** stands for Context-Aware preference mining by Tensor Factorization, which is a model-based CF approach introduced in [Karatzoglou et al. 2010]. In this approach, all users’ context logs can be represented by a high dimensional tensor $T$, and the objective is to complete the missing values for ranking content categories. Specifically, each value $t_{u,v_1,v_2,...,v_n}$ in original tensor is measured by counting how many context records of user $u$ contain contexts in category $c$ and feature value pairs $(p = (f_i, v_i))$.

At last, we select a naive approach which is context-aware but not personalized as the last baseline.

**CPP** stands for Context-aware Popularity based preference Prediction which is a basic context-aware recommendation approach without considering personal context-aware preference. To be specific, in this approach, given a user $u$ and a context $C$, we predict the user preferred content categories by the most frequent content categories appearing under $C$ according to all users’ historical context logs and then recommend corresponding content categories.

### 5.3. Evaluation Metrics

In the experiments, we utilize a five-fold cross validation to evaluate each test approach. To be specific, we first randomly divide each user’s context log into five equal parts, and then use each part as the test data while using other four parts as the training data in five test rounds. Finally, we report the average performance of the five runs. In the test process, we only take into account the context records with non-empty usage records, and use the contexts and the category of the content indicated by the
usage record as context inputs and ground truth, respectively. Moreover, to evaluate the ranking of content categories generated by each approach, we leverage two metrics as follows.

**MAP@K** stands for Mean Average Precision at top $K$ recommendation results. To be specific, $\text{MAP@K} = \frac{1}{|U|} \sum_{u} \sum_{r=1}^{K} (P_i(r) \times \text{rel}_i(r))$, where $N_u$ denotes the number of test cases for user $u$, $r$ denotes a given cut-off rank, $P_i(r)$ denotes the precision on the $i$-th test case of $u$ at a given cut-off rank $r$, and $\text{rel}_i(r)$ is the binary function on the relevance of a given rank.

**MAR@K** stands for Mean Average Recall at top $K$ prediction results. To be specific, $\text{MAR@K} = \frac{1}{|U|} \sum_{u} \sum_{r=1}^{K} \text{rel}_i(r)$.

### 5.4 Overall Results

According to the parameter estimation approaches introduced in Section 3 and Section 4, the number of common context-aware preferences for LDA, LDAC and NMF training denoted as $K$ is empirically set to be 15. The further robustness evaluation with vary $K$ will be shown in Section 5.5. For the LDA training, the two parameters $\alpha$ and $\beta$ are empirically set to be $50/K$ and 0.2 according to [Heinrich 2009]. The parameters for training LDAC model are similar to [Bao et al. 2010]. For the Bayesian Matrix Factorization training, according to [Schmidt 2009], we use an isotropic noise model and choose a decoupled prior for $\Phi$ and $\Theta$ with zero mean $\mu = 0$, and an unit diagonal covariance matrix $\Sigma = I$. The maximum iterations of Gibbs sampling are set to be 2000 in our experiments. Moreover, the behavior patterns are mined by BG-Growth algorithms introduced in [Li et al. 2012]. Both our approaches and the baselines are implemented by C++ and the experiments are conducted on a 3GHZ×4 quad-core CPU, 3G main memory PC.

Figure 8 shows the convergence curves of Gibbs sampling for our three implementations of the proposed approach by measuring their log likelihood for the training data set in one of the five test rounds. From the figure we can see that the Gibbs sampling of all implementations converge quickly. The convergence curves for other test rounds follow the similar trend. Moreover, each iteration of Gibbs sampling averagely costs 89 milliseconds for CIAP-LDA, 125 milliseconds for CIAP-LDAC and 423 milliseconds for CDAP, respectively. It is because that the NMF training is more complex than topic
models and the number of associations between context and content category for matrix factorization is greater than the ACP-features in both LDA and LDAC model.

It is worth noting that both our approach and baselines except CATF may be not able to generate recommendations for all contexts. To be specific, CIAP-LDA, CIAP-LDAC, CIAP-i, CRSVM, CACF can only generate preference predictions for the contexts which contain the ACP-features appeared in training sets, CDAP and CDAP-i can only generate preference predictions for the contexts which contain content-relevant sub-contexts, and CPP only can predict for the contexts under which there exist some activity records in training sets. In our experiments, we observe that CIAP-LDA, CIAP-LDAC, CIAP-i, CDAP and CRSVM can cover 100% test contexts while CDAP-i can averagely cover 89.45% test contexts and CPP can averagely cover 96.23% test contexts in all five round tests. In the evaluation of these approaches, we only count their MAP@K and MAR@K for the test cases whose contexts they cover.

We first test the MAP@K performance of each test approach with respect to varying K and the average results in the five-fold cross validation are shown in Figure 9 (a). From the experimental results we can get some insightful observations as follows. First, our approaches CDAP, CIAP-LDA and CIAP-LDAC consistently outperform other baselines with varying K, which may indicate our preference mining framework is more effective for mining context logs than other approaches, such as memory-based and model-based CF methods. Second, CIAP-LDAC outperforms CIAP-LDA slightly, which indicates the effectiveness of the contextual feature priori in LDAC for modeling ACP-feature bags. Third, we can see that CDAP outperforms CIAP-LDAC slightly with varying K, which validates that the context dependency assumption is more robust than the context conditional independency assumption. Fourth, the popularity based approach CPP always has the worst performance in predicting user preferences, which indicates that popularity based approach is not suitable for context-aware recommendation. Finally, we also find that two CF based approaches CACF and CATF outperform other individual based approaches (i.e., CRSVM, CIAP-i and CDAP-i), which indicates leveraging many users’ context logs other than individual users’ context logs can improve the recommendation performance. Figure 10 (a) and (b) further show the MAP@K of each recommendation approach in each test round when

\[ \text{MAP}_{@K} \]

\[ \text{MAR}_{@K} \]
Fig. 10. The $MAP@K$ of each prediction approach when (a) $K = 5$, (b) $K = 8$ in each test round.

Fig. 11. The $MAR@K$ of each prediction approach when (a) $K = 5$, (b) $K = 8$ in each test round.

$K = 5$ and $K = 8$, respectively. From the results we can observe that our approaches consistently outperform other baselines in all five test rounds.

Figure 9 (b) shows the average $MAR@K$ of each approach in five-fold cross validation. From the results, we can observe it has the similar trends as $MAP@K$ performance. Specifically, we find that CIAP-LDA, CIAP-LDAC and CDAP outperform other baselines and CDAP outperforms CIAP-LDAC slightly in terms of $MAR@K$. Figure 11 (a) and (b) further show the $MAR@5$ and $MAR@8$ of each recommendation approach in each test round. The results also validate the effectiveness of our approaches in terms of $MAR@K$ performance.

Specially, we conduct a series of paired T-tests with 0.95 confidence level in each $K$. The test results show that the improvements of CIAP-LDA, CIAP-LDAC and CDAP on $MAP$ and $MAR$ compared with other baselines are statistically significant. We also study the variances of overall $MAP@K$ and $MAR@K$ of all tested approaches in the five-fold cross validation for validating the effectiveness of experimental results. Table IV shows the mean deviations of these values of each tested approach in the five-fold cross validation with $K = 1$, $K = 5$ and $K = 8$. From this table we can observe that the variances of our approaches (i.e., CIAP-LDA, CIAP-LDAC, CDAP) and the two CF based approaches (i.e., CACF, CATF) consistently smaller than other baselines. It implies that individual user context logs are often very limited, which may influence the stability of recommendation performance.

From the above experimental results we can clearly see that both CDAP, CIAP-LDA, CIAP-LDAC outperform other baselines under different metrics and experimental settings, which demonstrates the effectiveness of our framework for personalized
context-aware recommendation. Moreover, CDAP outperforms CIAP-LDA and CIAP-LDAC slightly though its training cost is much higher than the latter.

5.5. Robustness Analysis

Since CIAP-LDA, CIAP-LDAC and CDAP need a parameter to determine the number of common context-aware preferences (i.e., $K$). Figure 12 (a) and (b) show the $MAP@5$ and $MAP@10$ of CIAP-LDA, CIAP-LDAC and CDAP with respect to varying settings of the number. From these figures we can observe that both $MAP@5$ and $MAP@10$ of CDAP are relatively not sensitive to the parameter. In contrast, the robustness of CIAP-LDA and CIAP-LDAC are not good with small numbers of common context-aware preferences but becomes stable when the setting of the number increases. It may be because that CDAP leverages associations between contexts and user content categories for extracting common context-aware preferences and such associations have been filtered from noisy data. Thus, the quality of mined common context-aware preferences is always relatively good with different parameters since the mining are on the basis of pruned training data. In contrast, CIAP-LDA and CIAP-LDAC leverage ACP-features for extracting common context-aware preferences, where ACP-features usually contain more noisy information and thus make the mining results more sensitive to parameters.

5.6. Case Study

In addition to the studies on the overall performance of our approach, we also study the cases in which CIAP-LDA, CIAP-LDAC and CDAP outperform the baselines.

Table IV. The results of the mean deviations of $MAP@K$ and $MAR@K$ of each tested approach in the 5-fold cross validation.

<table>
<thead>
<tr>
<th>Approach</th>
<th>$MAP@1^*$</th>
<th>$MAP@5$</th>
<th>$MAP@8$</th>
<th>$MAR@5$</th>
<th>$MAR@8$</th>
</tr>
</thead>
<tbody>
<tr>
<td>CIAP-LDA</td>
<td>$5.73 \times 10^{-3}$</td>
<td>$5.54 \times 10^{-3}$</td>
<td>$2.96 \times 10^{-3}$</td>
<td>$1.32 \times 10^{-3}$</td>
<td>$6.35 \times 10^{-3}$</td>
</tr>
<tr>
<td>CIAP-LDAC</td>
<td>$8.32 \times 10^{-3}$</td>
<td>$4.13 \times 10^{-3}$</td>
<td>$2.90 \times 10^{-3}$</td>
<td>$1.46 \times 10^{-3}$</td>
<td>$5.97 \times 10^{-3}$</td>
</tr>
<tr>
<td>CDAP</td>
<td>$7.29 \times 10^{-3}$</td>
<td>$3.87 \times 10^{-3}$</td>
<td>$1.04 \times 10^{-3}$</td>
<td>$8.23 \times 10^{-4}$</td>
<td>$4.21 \times 10^{-3}$</td>
</tr>
<tr>
<td>CIAP-i</td>
<td>$1.52 \times 10^{-2}$</td>
<td>$9.22 \times 10^{-3}$</td>
<td>$4.15 \times 10^{-3}$</td>
<td>$4.43 \times 10^{-2}$</td>
<td>$8.72 \times 10^{-3}$</td>
</tr>
<tr>
<td>CDAP-i</td>
<td>$1.31 \times 10^{-2}$</td>
<td>$8.59 \times 10^{-3}$</td>
<td>$3.87 \times 10^{-3}$</td>
<td>$4.02 \times 10^{-2}$</td>
<td>$9.11 \times 10^{-3}$</td>
</tr>
<tr>
<td>CASVM</td>
<td>$2.84 \times 10^{-2}$</td>
<td>$1.34 \times 10^{-2}$</td>
<td>$5.44 \times 10^{-3}$</td>
<td>$5.52 \times 10^{-2}$</td>
<td>$2.24 \times 10^{-2}$</td>
</tr>
<tr>
<td>CACF</td>
<td>$9.97 \times 10^{-3}$</td>
<td>$7.01 \times 10^{-3}$</td>
<td>$3.38 \times 10^{-3}$</td>
<td>$2.38 \times 10^{-2}$</td>
<td>$7.98 \times 10^{-3}$</td>
</tr>
<tr>
<td>CATF</td>
<td>$1.27 \times 10^{-2}$</td>
<td>$8.22 \times 10^{-3}$</td>
<td>$3.61 \times 10^{-3}$</td>
<td>$1.74 \times 10^{-2}$</td>
<td>$7.35 \times 10^{-3}$</td>
</tr>
<tr>
<td>CPP</td>
<td>$4.82 \times 10^{-2}$</td>
<td>$2.94 \times 10^{-2}$</td>
<td>$5.89 \times 10^{-3}$</td>
<td>$7.38 \times 10^{-3}$</td>
<td>$4.02 \times 10^{-3}$</td>
</tr>
</tbody>
</table>

$MAP@1$ and $MAR@1$ have same value for each tested approach.

Fig. 12. The (a) $MAP@5$ and (b) $MAP@10$ of CIAP and CDAP with respect to varying number of common context-aware preferences.

ACM Transactions on Intelligent Systems and Technology, Vol. V, No. N, Article A, Publication date: January YYYY.
For example, Table V shows an example of the top 3 predicted preferences of each approach for two different users given the same context 
{(Time range: 21:00-22:00), (Profile: Silent), (Is holiday: Yes), (Day name: Sunday), (Day period: Night), (Location: Home)}, which may imply the leisure time at home. Under that context, the ground truth preferred content categories of user #162 and user #423 are Game and Web, re-
Mining Mobile User Preferences...

Table VII. The common preference \( z^* \) mined by CIAP-LDA for user #162.

<table>
<thead>
<tr>
<th>preference</th>
<th>content category</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>(Is holiday?: Yes) → Game</td>
</tr>
<tr>
<td>2</td>
<td>(Is holiday?: Yes) → Multimedia</td>
</tr>
<tr>
<td>3</td>
<td>(Day period: Evening) → Game</td>
</tr>
<tr>
<td>4</td>
<td>(Location: Home) → Web</td>
</tr>
<tr>
<td>5</td>
<td>(Location: Home) → Game</td>
</tr>
<tr>
<td>6</td>
<td>(Profile: Silent) → Game</td>
</tr>
<tr>
<td>7</td>
<td>(Profile: General) → Web</td>
</tr>
<tr>
<td>8</td>
<td>(Location: Home) → Multimedia</td>
</tr>
<tr>
<td>9</td>
<td>(Day period: Evening) → Web</td>
</tr>
<tr>
<td>10</td>
<td>(Profile: Silent) → Game</td>
</tr>
</tbody>
</table>

Table VIII. The common preference \( z^* \) mined by CIAP-LDAC for user #162.

<table>
<thead>
<tr>
<th>preference</th>
<th>content category</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>(Is holiday?: Yes) → Game</td>
</tr>
<tr>
<td>2</td>
<td>(Location: Home) → Game</td>
</tr>
<tr>
<td>3</td>
<td>(Is holiday?: No) → Web</td>
</tr>
<tr>
<td>4</td>
<td>(Location: Home) → Web</td>
</tr>
<tr>
<td>5</td>
<td>(Profile: Silent) → Game</td>
</tr>
<tr>
<td>6</td>
<td>(Day period: Night) → Game</td>
</tr>
<tr>
<td>7</td>
<td>(Profile: General) → Web</td>
</tr>
<tr>
<td>8</td>
<td>(Location: Home) → Web</td>
</tr>
<tr>
<td>9</td>
<td>(Day name: Sunday) → Web</td>
</tr>
<tr>
<td>10</td>
<td>(Location: On the way) → Web</td>
</tr>
</tbody>
</table>

Table IX. The common preference \( z^* \) mined by CDAP for user #162.

<table>
<thead>
<tr>
<th>preference</th>
<th>content category</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>(Is holiday?: Yes) → Game</td>
</tr>
<tr>
<td>2</td>
<td>(Location: Home) → Web</td>
</tr>
<tr>
<td>3</td>
<td>(Is holiday?: Yes) → Multimedia</td>
</tr>
<tr>
<td>4</td>
<td>(Profile: General) → Web</td>
</tr>
<tr>
<td>5</td>
<td>(Location: Home) → Game</td>
</tr>
<tr>
<td>6</td>
<td>(Profile: Silent) → Game</td>
</tr>
<tr>
<td>7</td>
<td>(Is holiday?: No) → Web</td>
</tr>
<tr>
<td>8</td>
<td>(Profile: General) → Web</td>
</tr>
<tr>
<td>9</td>
<td>(Day name: Sunday) → Web</td>
</tr>
<tr>
<td>10</td>
<td>(Location: On the way) → Web</td>
</tr>
</tbody>
</table>

spectively. From the results, we can observe that CIAP-LDA, CIAP-LDAC and CDAP can predict correct content categories for user #162 and user #423 in the top one position. In contrast, CIAP-i, CDAP-i, CASVM, CACF and CATF predicted correct content categories in lower positions for both users. Moreover, the popularity based approach CPP always predicts the same preferences for both users and thus performs not well.

Furthermore, Table VI shows another example of the top 3 predicted preferences of each approach for user #162 and user #423. The given context is \( \{ \) (Time range: PM18:00-19:00), (Profile: General), (Is holiday: No), (Day name: Monday), (Day period: Evening), (Location: on the way) \( \) \}, which is similar to the context introduced in our motivating example (i.e., Example 1.1). In this case, the ground truth content categories of user #162 and user #423 are Web and Multimedia, respectively. From the results we can observe the similar trend of predicting performance with last case.

Indeed, our approaches outperform other baselines, especially the two CF based approaches (i.e, CACF, CATF), is because that our mining framework can represent personal context-aware preferences of mobile users in a more reasonable way, which can model contexts with respect to different data dependency assumptions. Thus, the three approaches developed based on our framework can be more effective for preference predicting by mining context logs. Moreover, the individual based approaches (i.e., CR SVM, CIAP-i and CDAP-i) perform worse in preference predicting than our approaches is because that individual user’ context logs are often very limited, which may influence the stability of recommendation performance.

Finally, to further study the reason why our approaches can outperform other baselines, we investigate the mined personal context-aware preferences for user #162. To be specific, in our framework, the personal preference of each user can be represented as a distribution of these common context-aware preferences. In fact, the common context-aware preference \( z^* \) that a user \( u \) has highest probability of belonging to, i.e., \( z^* = \arg \max_z P(z|u) \), will have the most influential impact on the personal preference of \( u \). Therefore, for the user #162, we manually inspect the most important common preference \( z^* \), and study whether it contains some relationships between the given context and ground truth content. Table VII to Table IX show the different \( z^* \) mined

ACM Transactions on Intelligent Systems and Technology, Vol. V, No. N, Article A, Publication date: January YYYY.
by CIAP-LDA, CIAP-LDAC and CDAP for user #162, respectively. Note that, the common preferences mined by both CIAP-LDA and CIAP-LDA are constituted by ACP-features, while the the common preferences mined by CDAP are constituted by behavior patterns (limited to space, we only show top 10 ACP-features and behavior patterns in each common preference). From these tables, we can observe that there are many ACP-features and behavior patterns (i.e., labeled in bold) have explicit relationships with the given context and ground truth content category. Therefore, our approaches can predict the corresponding content category with high ranking in the results. As a conclusion, the above case studies clearly validate the effectiveness of our preference mining framework and approaches.

6. RELATED WORK

Many previous works about personalized context-aware recommendation for mobile users have been reported. For example, Tung et al. [Tung and Soo 2004] have proposed a prototype design for building a personalized recommender system to recommend travel related information according to users’ contextual information. Park et al. [Park et al. 2007] proposed a location-based personalized recommender system, which can reflect users’ personal preferences by modeling user contextual information through Bayesian Networks. Bader et al. [Bader et al. 2011] have proposed a novel context-aware approach to recommending points-of-interest (POI) for users in an automotive scenario. Specifically, they studied the scenario of recommending gas stations for car drivers by leveraging a Multi-Criteria Decision Making (MCDM) based methods to modeling context and different routes. However, most of these works only leverage individual users’ historical context data for modeling personal context-aware preferences, and do not take into account the problem of insufficient personal training data.

Actually, the problem of insufficient personal training data is common in practice and many researchers have studied how to address this problem. For example, Woerndl et al. [Woerndl et al. 2007] proposed a hybrid framework named “play.tools” for recommending mobile applications by leveraging users’ context information. This recommendation framework are based on what other users have installed in similar context will be liked by a given user. Kim et al. [jae Kim et al. 2010] investigated several Collaborative Filtering (CF) based approaches for recommendation and developed a memory based CF approach to providing context-aware advertisement recommendation. Specially, the proposed approach can leverage a classification rule of decision tree to understand users’ personal preference. Zheng et al. [Zheng et al. 2010] have studied a model based CF approach to recommending user locations and activities according to users’ GPS trajectories. The approach can model user, location and activity as a 3-dimensional matrix, namely tensor, and perform tensor factorization with several constraints to capture users’ preferences. Alexandros et al [Karatzoglou et al. 2010] proposed a model based CF approach for making recommendation with respect to rich contextual information, namely multiverse recommendation. Specifically, they modeled the rich contextual information with item by N-dimensional tensor, and proposed a novel algorithm to make tensor factorization. In a word, most of these approaches are based on rating logs of mobile users and the objective is to predict accurate ratings for the unobserved items under different contexts. However, we usually cannot obtain such rating data in user mobile devices. In contrast, it is relatively easier to collect context logs which contain the users’ historical context information and their usage records, which can be used for mining context-aware user preferences.

Recently, some researchers studied how to mine event logs for personalized context-aware recommendation. For example, Lee et al [Lee et al. 2010] investigated to convert the event logs into implicit ratings and tested various memory based CF approaches to make personalized context-aware recommendation. Kahng et al [Kahng et al. 2011]
proposed a novel approach for ranking item in event logs for personalized context-aware recommendation. Compared with context logs, event logs only record the context records with non-empty usage records and lose thus some discriminative information to capture the relevance between contexts and content categories [Cao et al. 2010]. But these event log based approaches can still be used for mining context logs towards personalized context-aware recommendation. Recently, Yu et al. [Yu et al. 2012] proposed a novel personalized context-aware mobile recommender system by analyzing mobile user’s context logs. The proposed approach is based on Latent Dirichlet Allocation topic model and scalable for multiple contextual features. However, this approach only focuses on the situation of context independent assumption, which is similar to our CIAP-LDA approach. Different with these works, we propose a novel framework which can explicitly model common context-aware preferences and represent individual users’ personal context-aware preferences by distributions of common context-aware preferences with different context dependency assumptions. Moreover, one proposed approach in our framework (CDAP) can utilize the context records with only empty usage records and context dependent assumption.

7. CONCLUDING REMARKS

In this paper, we proposed to exploit user context logs for mining the personal context-aware preferences of mobile users. First, we identified common context-aware preferences from the context logs of many users. Then, the personal context-aware preference of an individual user can be represented as a distribution of common context-aware preferences. Moreover, we designed two methods for mining common context-aware preferences based on two different assumptions about context data dependency. Finally, the experimental results on a real-world data set clearly showed that the proposed approach could achieve better performances than benchmark methods for mining personal context-aware preferences, and the one implementation based on the independent assumption of context data slightly outperforms another one but has higher computational cost.

As mentioned above, the proposed user context-aware preference mining approaches are based on the analysis of many users’ daily context logs, which will be very huge in quantity. Therefore, how to efficiently store and exploit our mining approaches on such “big data” are very challenging problems to be further studied. Moreover, how to capture users’ preference drifts and protect their privacy are also valuable research aspects in our future plan. Last but not least, in the future we will try to integrate our approaches with various real-world services, such as mobile context-aware advertising, for enhancing user experience.

Acknowledgement. This work was supported in part by grants from Natural Science Foundation of China (NSFC, Grant No. 61073110 and 71329201), China National Funds for Distinguished Young Scientists, Research Fund for the Doctoral Program of Higher Education of China (Grant No. 20113402110024), the Key Program of National Natural Science Foundation of China (Grant No. 60933013), and National Key Technology Research and Development Program of the Ministry of Science and Technology of China (Grant No. 2012BAH17B03). This work was also partially supported by grants from National Science Foundation (NSF) via grant numbers CCF-1018151 and IIS-1256016.

REFERENCES


ACM Transactions on Intelligent Systems and Technology, Vol. V, No. N, Article A, Publication date: January YYYY.
A. APPENDIX: GIBBS SAMPLING DERIVATION FOR CIAP-LDAC APPROACH

In this paper, we leverage the Gibbs sampling based approach introduced in [Bao et al. 2010] to train LDAC model in our CIAP-LDAC approach. To be specific, this method begins with a random assignment of common context-aware preferences to ACP-features for initializing the state of Markov chain. In each of the following iterations, the method will re-estimate the conditional probability of assigning a common context-aware preference to each ACP-feature, which is conditional on the assignment of all other ACP-features. Then a new assignment of common context-aware preferences to ACP-features according to those latest calculated conditional probabilities will be scored as a new state of Markov chain. Finally, after rounds of iterations, the assignment will converge, which means each ACP-feature is assigned a stable and final common context-aware preference.
According to the introduction in Section 3.1.2 and the graphical representation in Figure 4 (b), we have

\[
P(d, \theta, z_d, \pi_d, \Phi|\alpha, \beta, \gamma) = P(\theta_d|\alpha)P(\Phi|\beta)P(\pi_d|\gamma)
\times \prod_{i=1}^{N_d} P(c_i, v_{d,i}|z_{d,i}, f_{d,i}, \Phi)P(f_{d,i}|\pi_d)P(z_{d,i}|\theta_d),
\]

where \( \Phi = \{\phi_{k,f}\}, z_d = \{z_{d,i}\} \), and \( N_d \) denotes the number of ACP-features in the ACP-feature bag \( d \). If we denote the token \((d, i)\) as \( m \), the Gibbs sampler of common context-aware preference \( z_m \) in each sampling round can be computed as follows.

\[
P(z_m = k|Z_{\sim m}, D) \propto P(z_m = k, Z_{\sim m}, D)
\times P(c_i, v_m|z_m = k, Z_{\sim m}, F, V_{\sim m})
\times P(z_m = k|Z_{\sim m}),
\]

where \( \sim m \) means removing the contextual feature-value pair \((f_m : v_m)\) from corpus \( D \), and \( Z \) is the labels of the common context-aware preferences in \( D \). Furthermore, we have the following estimation

\[
P(c_i, v_m|z_m = k, Z_{\sim m}, F, V_{\sim m}) = \sum_v \frac{n_{\sim m,k,f_m,c_i,v} + \beta_{c_i,v}}{\sum_{v'} n_{\sim m,k,f_m,c_i,v'} + \sum_{v'' \in V_{f_m}} \beta_{v''}}
\]

\[
P(z_m = k|Z_{\sim m}) = \frac{n_{\sim m,k} + \alpha_k}{\sum_{k'=1}^{K} n_{\sim m,k'} + \sum_{k'=1}^{K} \alpha_{k'}}
\]

where \( n_{\sim m,k,f,c,v} \) indicates the frequency that the ACP-feature \((c, p) (p = (f : v))\) is labeled with the \( k \)-th common context-aware preference in all ACP-feature bags after removing the \( m \)-th ACP-feature, and \( n_{d,\sim m,k} \) indicates the number of ACP-features labeled with the \( k \)-th common context-aware preferences in \( d \) expect for the \( m \)-th one.

After training process, we can obtain the probabilities \( P(z|u) \) by Equation 1 and \( P(c, p|z) \) as follows.

\[
P(c, p|z_k) = P(c, v_p|f_p, z)P(f_p),
\]

where \( p = (f_p : v_p) \) and

\[
P(c, v_p|f_p, z) = \frac{n_{z,c,f_p,v_p} + \beta_{c,v_p}}{\sum_{v} n_{z,c,f_p,v} + \sum_{v'' \in V_{f_m}} \beta_{v''}}
\]

\[
P(f_p) = \frac{\sum_{z} \sum_{v} n_{z,c,f_p,v} + \gamma_{f_m}}{\sum_{f} \sum_{z} \sum_{v} n_{z,c,f_p,v} + \sum_{f} \gamma_{f}}
\]

where \( n_{z,c,f_p,v} \) is the frequency of ACP-feature \((c, p)\) has been assigned to common context-aware preference \( z \), and \( V_{f_m} \) is the number of values for contextual feature \( f_m \).