A Habit Mining Approach for Discovering Similar Mobile Users

Haiping Ma[1,2]   Huanhuan Cao[2]   Qiang Yang[3]   Enhong Chen[1]   Jilei Tian[2]

[1]University of Science and Technology of China, No.96 JinZhai Road, Hefei, Anhui
[2]Nokia Research Center GEL, No.5 Donghuan Zhonglu, BDA, Beijing 100176
[3]Hong Kong University of Science and Technology, Clearwater Bay, Kowloon, Hong Kong
[1]{mhp0814, cheneh}@ustc.edu.cn   [2]{happia.cao, jilei.tian}@nokia.com   [3]qyang@cse.ust.hk

ABSTRACT

Discovering similar users with respect to their habits plays an important role in a wide range of applications, such as collaborative filtering for recommendation, user segmentation for market analysis, etc. Recently, the progressing ability to sense user contexts of smart mobile devices makes it possible to discover mobile users with similar habits by mining their habits from their mobile devices. However, though some researchers have proposed effective methods for mining user habits such as *behavior pattern mining*, how to leverage the mined results for discovering similar users remains less explored. To this end, we propose a novel approach for conquering the sparseness of behavior pattern space and thus make it possible to discover similar mobile users with respect to their habits by leveraging behavior pattern mining. To be specific, first, we normalize the raw context log of each user by transforming the location-based context data and user interaction records to more general representations. Second, we take advantage of a constraint-based Bayesian Matrix Factorization model for extracting the latent common habits among behavior patterns and then transforming behavior pattern vectors to the vectors of mined common habits which are in a much more dense space. The experiments conducted on real data sets show that our approach outperforms three baselines in terms of the effectiveness of discovering similar mobile users with respect to their habits.

Categories and Subject Descriptors

H.2.8[Database Management] [**Database Applications**]: Data Mining; I.2.6[Artificial Intelligence] [**Learning**]

General Terms

Algorithms, Experimentation

Keywords

Discover similar users, habit mining, mobile users

1. INTRODUCTION

Discovering similar users with respect to their habits plays an important role in a wide range of applications such as collaborative filtering for recommendation [7, 26, 25], user

segmentation for market analysis [27, 21], etc. Although the approach for capturing and comparing the similarity of different users' habits has been well studied in the field of social science [12], most of them are too expensive to be applied to crowd users in practice. Compared with the poor scalability of traditional approaches, in the last decade, Web-based user habit mining and comparison attracted lots of attention due to the relatively low cost. For example, Mobasher et al. [16] proposed and experimentally evaluated two technologies in order to mine user habits by discovering overlapping aggregate profiles. Zhou et al. [27] proposed to analyze the online behaviors of Web users and segment them through a mixture of factor analyzers.

In recent years, the increasing popularity of smart mobile devices and their more and more powerful ability of sensing user contexts make it possible to capture rich contextual information and user interactions through context logs (e.g., Table 1 shows a toy context log) in devices. Context logs can be used for mining the habits of mobile users and then discovering similar users with respect to their habits. Compared with Web-based user habit mining approaches, the major advantage of mobile habit mining is that it can capture richer contextual information of users while capturing user activity information due to the more powerful sensors of smart mobile devices, which implies the possibility of characterizing user habits more precisely. The following example intuitively illustrates the advantage of mobile habit mining in terms of capturing user habits.

EXAMPLE 1   (A MOTIVATING EXAMPLE). *Suppose that Sam is a fan of the social game FarmVille and he usually plays the game through his PC at home in nights and through his smart phone in the bus leaving from his work place in evenings, a Web-based habit mining approach may discover Sam usually "plays games in nights". In contrast, a mobile habit mining approach may discover Sam usually "plays games in the bus (sensing by 3D accelerometer) from his work place (sensing by GPS or cell ID sensor combined by social location mining technology) to home in evenings", which is more precise for characterizing the habit of Sam.*

Cao et al. [3] proposed an effective approach for mobile habit mining. They mined associations between interactions and contexts, i.e., *behavior patterns*, from the context logs of mobile users for representing user habits and gave an effective algorithm for solving the problem. By using different measurements of support and confidence, they successfully addressed the unbalanced occurrences of context data and interaction records which challenge the traditional approaches for association rule mining [1, 8].

Table 1: A toy context log.

| Timestamp	Context	Interaction record
t_1	{(Is a holiday?: No),(Time range: AM8:00-9:00),(Cell ID: 2341-42344)}	Null
t_2	{(Is a holiday?: No),(Time range: AM8:00-9:00),(Cell ID: 2341-22347)}	Playing Music Player
t_3	{(Is a holiday?: No),(Time range: AM8:00-9:00),(Cell ID: 2341-79901)}	Null
	......	
t_{38}	{(Is a holiday?: No),(Time range: AM8:00-9:00),(Cell ID: 2341-32044)}	Null
t_{39}	{(Is a holiday?: No),(Time range: AM8:00-9:00),(Cell ID: 2341-2501)}	Null
	......	
t_{58}	{(Is a holiday?: No),(Time range: AM8:00-9:00),(Cell ID: 2341-42344)}	Playing Music Player
t_{59}	{(Is a holiday?: No),(Time range: AM8:00-9:00),(Cell ID: 2341-42344)}	Null
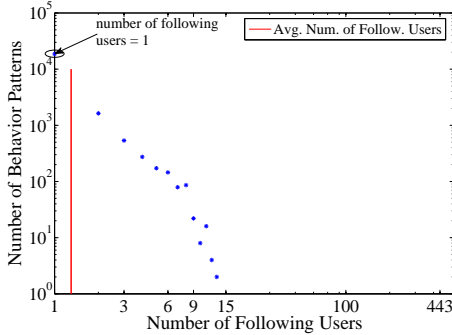
Figure 1: The distribution of mined behavior patterns with respect to the numbers of their following users in a real data set.

Intuitively, we can first map users into the behavior pattern space by their behavior patterns and the corresponding confidences and then find nearby data points in the space for discovering similar users with respect to their habits. However, this naive approach does not work in practice because we observe that the behavior pattern space for crowd users is usually very sparse. For example, Figure 1 illustrates the distribution of mined behavior patterns with respect to the numbers of their *following users*[1] in a real data set which includes 443 users' context logs. From the figure we can see that most behavior patterns are followed infrequently, which implies that the behavior pattern space is extremely sparse. As is well known, there is no distance metric which can work well in an extremely high-dimensional and sparse space, which is referred to as *the curse of dimensionality* [2].

To this end, motivated by some observations from real data, we propose a two-stage approach for addressing the sparseness of behavior pattern space. To be specific, first of all, we normalize the raw context logs by transforming the location-based context data to social locations and transforming the raw interaction records into interaction categories which are more general and able to reflect the common points of different interactions. Second, we further take advantage of the Matrix Factorization technology for extracting common habits of users denoted as *hyper behavior patterns* from the matrix of mined behavior patterns and users. Finally, we map users into the space of hyper behavior patterns which has much less dimensions and is much more dense. To incorporate some particular constraints in our problem, we use a constraint-based Bayesian Matrix Factorization model [17], which can naturally address the Matrix Factorization problems with some particular inequality or e-

---

[1]If a user u follows a behavior pattern p, we refer to u as a following user of p.

quality constraints. We also conduct extensive experiments on real data for evaluating the proposed approach. The experimental results clearly show that the approach outperforms the baselines in terms of both the ability of reducing the sparseness of behavior pattern space and the effectiveness of discovering similar users with respect to their habits.

The remainder of the paper is organized as follows. In Section 2, we briefly introduce some preliminaries and outline the general framework of our approach. Section 3 introduces how we normalize raw context logs by transforming location-based context data and raw interaction records. Followed in Section 4, we present the details of how we take advantage of Matrix Factorization technology for extracting hyper behavior patterns. Next, we report and analyze the experimental results in Section 5 and briefly discuss some related work in Section 6. Finally, we conclude this paper in Section 7 and look forward the direction of future work.

2. OVERVIEW

2.1 Preliminaries

The problem of behavior pattern mining was proposed by Cao et al. [3] for mining the habits of mobile users from their context logs which record their historical context data and interaction records. Table 1 shows a toy context log which contains several context records. Each context record consists of a timestamp, the most detailed available context at that time, and a corresponding user interaction record. We mention "available" because a context record may miss some context data though the context data to be collected are usually predefined. For example, the GPS coordinate is not available when the user stays indoors. Moreover, interaction records can be empty (denoted as "Null") because user interactions do not always happen.

According to [3], the associations between user contexts and the user interactions with mobile devices are regarded as behavior patterns if both their supports and confidences are bigger than the predefined thresholds. Compared with the traditional association rule mining, behavior pattern mining can address the unbalanced occurrences of context data and user interaction records well due to its different way of calculating the supports of contexts. To be specific, in behavior pattern mining the support of a context is counted by taking into account how many times it continuously appears in several adjacent context records, which constitute a *context range* of the context. For a context range which contains non-empty interaction records, the number of non-empty interaction records is regarded as the support of the corresponding context in this context range. Otherwise, for a context range which only contains empty interaction records, the support of the corresponding context is regarded as one in this context range. Finally, the support of the cor-

responding context is calculated by summing the supports in each of its context ranges. Take the context log in Table 1 for example, let's consider the context "$\{$*(Is a holiday?: No),*(*Time range: AM8:00-9:00*)$\}$". Though it appears in seven context records, its support is not seven but three because it has three context ranges, i.e., (t_1, t_2, t_3), (t_{38}, t_{39}), and (t_{58}, t_{59}), respectively, and each context range has at most one non-empty interaction record. For more details of behavior pattern mining, please refer to [3].

2.2 Addressing the Sparseness of Behavior Pattern Space

Intuitively, to find similar users by leveraging behavior patterns, we can first build behavior pattern vectors for users by considering their behavior patterns and the corresponding confidences. Then we can calculate the similarity between any two users by calculating the distance between their behavior pattern vectors. However, this naive approach does not work in practice because we observe that the behavior pattern space for crowd users is usually extremely sparse (e.g., Figure 1 in Section 1 illustrates the sparseness of the mined behavior patterns from a real data set), which makes it difficult to find similar users by leveraging behavior patterns. We observe that this phenomenon is majorally caused by two problems as follows.

The first problem is that in raw context logs the location-based context data and user interaction records are too specific to reflect the common points among different users' behavior patterns. For example, given two behavior patterns mined from two different users' context logs "$\{$*(Is a holiday?: No),*(*Time range: PM8:00-9:00),*(*Cell ID: 2341-2344*)$\}\longrightarrow$ *Playing Angry Birds*" and "$\{$*(Is a holiday?: No),*(*Time range: PM8:00-9:00),*(*Cell ID: 25001-8344*)$\}\longrightarrow$ *Playing Fruit Ninja*", at the first glance they are different. However, if we have known "$\{$*(Cell ID: 2344-2341*)$\}$" and "$\{$*(Cell ID: 25001-8344*)$\}$" indicate the homes of the first and second users, it may be not reasonable to regard the two behavior patterns totally different. Moreover, if we further know that both "Fruit Ninja" and "Angry Birds" are popular action games with similar features (e.g., cartoon figures and touch based operations), it is reasonable to assume that the two users have very similar habits.

The second problem is that there exist many behavior patterns which have different representations but indirect relationships. For example, given two behavior patterns mined from two different users' context logs "$\{$*(Is a holiday?: No),*(*Time range: PM8:00-9:00),*(*Cell ID: 2341-2344*)$\}\longrightarrow$ *Playing Angry Birds*" and "$\{$*(Is a holiday?: No),*(*Time range: PM7:00-8:00),*(*Cell ID: 2341-2344*)$\}\longrightarrow$ *Playing Angry Birds*", their contexts seem similar since both "*(Time range: PM8:00-9:00)*" and "*(Time range: PM7:00-8:00)*" imply evenings. If we can capture the latent common points between such behavior patterns, this problem may be addressed.

Motivated by above observations, we propose a novel approach for addressing the sparseness of behavior pattern space which includes two stages as follows. In the first stage, we normalize the raw context log of individual users. To be specific, we firstly transform location-based raw context data to social locations such as "Home" and "Work Place". Then we transform raw user interaction records to interaction categories such as "Email", "Action Games" and "Social Network Services". Thus, we effectively address the first problem listed above. In the second stage, we take advan-
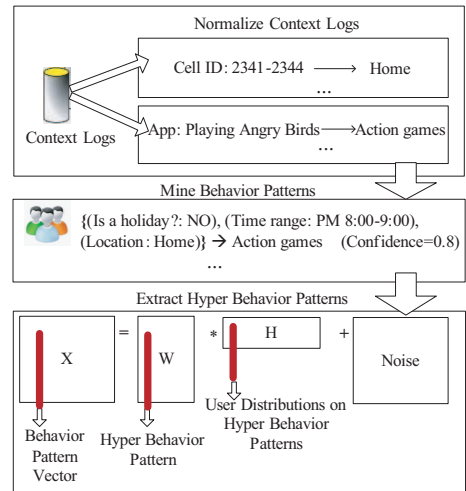

Figure 2: The framework of our approach.

tage of the Matrix Factorization technology for extracting the latent habits reflected by multiple unique behavior patterns, which are referred to as *hyper behavior patterns*, and represent each user as a mixture of hyper behavior patterns. Since the space of hyper behavior patterns is much more dense than the original behavior pattern space, it is easier to calculate the similarity between different users with respect to their habits in the new space. Figure 2 illustrates the overview of our approach. More details of the two stages are presented in the following sections.

3. NORMALIZING CONTEXT LOGS

3.1 Normalizing Location-based Context

In raw context logs, the location-based context data include the cell IDs which indicate the IDs of serving sell-cites and the geo-coordinates captured by GPS sensors. Among these data, geo-coordinates do not appear in each context record because the GPS signal is not available when the user stays indoors and the GPS positioning module may be disabled in order to save the battery life. As mentioned before, the behavior patterns which contain these data are too particular to be used for comparing different users' habits because different location-based raw context data may have the same social meaning such as "Home" and "Work Place". If we can transform the location-based context data in raw context logs to social locations, we may be able to improve the comparability of the mined behavior patterns for different users. Along this line, we need to define a list of social locations to be transformed. Ideally, it is desirable to define a social location list as comprehensively as possible to cover all social locations appearing in context logs. However in reality, it is difficult to accurately recognize all social locations of mobile users from their location-based context data. It is because the geo-coordinates are usually not available and we can only take advantage of cell IDs which are very imprecise for positioning. Moreover, even geo-coordinates are not always precise enough to indicate a social location. For instance, suppose that a geo-coordinate (longitude=x, latitude=y) indicates a position which is both near to a "Pub" and a "Hotel", it is difficult to decide which is the real social location of the user. To this end, an alternative approach is

just defining some social locations which can be recognized by only cell IDs for most users and omitting others. To be specific, we can transform cell IDs into "Home", "Work Place" and "Others" by taking advantage of the users' cell ID trajectories which consist of cell IDs ordered in timestamps and can be easily extracted from raw context logs.

Mining social locations from cell ID trajectories has been well studied by many researchers (e.g., [23, 13]). In this paper, we use a commonly used approach proposed by Yang [23]. The basic idea of the approach is as follows. First, it tries to discover the *closure patterns* whose the first and the last cell ID are duplicated while the number of unique cell IDs is less than a predefined threshold. Second, it iteratively merges discovered closure patterns which have common cell IDs. Third, each final closure pattern is regarded as a meaningful social location under time pattern analysis. Specially, for each closure pattern, we draw its normalized time histogram as illustrated by Figure 3. For a closure pattern P, the time ratio in time range t_i is calculated as $\frac{\sum_{c \in P} freq_{c,t_i}}{\sum_t \sum_{c \in P} freq_{c,t}}$, where $freq_{c,t}$ denotes the frequency of cell ID c appearing in the time range t. If the time ratio of P between AM8:00-PM6:00 is larger than a predefined threshold α, it is regarded as a candidate for "Work Place". Otherwise, if the time ratio of P in other time ranges is bigger than a predefined threshold β, it is regarded as a candidate for "Home". Finally, in all candidates for "Work Place" the closure pattern with the maximum total frequency is regarded as the final "Work Place". The selection of final "Home" is performed in a similar way.
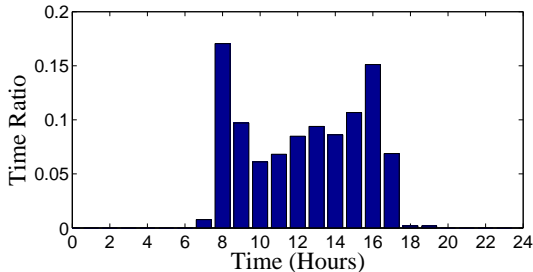

Figure 3: The normalized time histogram of an exemplary cell ID closure pattern.

3.2 Normalizing User Interaction Records

After normalizing the location-based contexts, we can further normalize user interaction records for extracting the common points of different users' activities. To be specific, we define a set of interaction categories which are more general to reflect user activities and classify each unique interaction record into a interaction category. Though we can manually label the interaction records for a small data set, it may be expensive for manually classifying interaction records for large scale data sets in practice. Therefore, we take advantage of a Web knowledge based automatic approach for classifying interaction records which is introduced in [4]. Firstly, we manually select some typical interaction records for each interaction category as *category seeds*. Then, we take advantage of search engines to obtain top 10 search results for each category seed and extract their snippets as textual descriptions of the corresponding category seeds. Thirdly, we remove all stop words such as "of", "the" from the extracted snippets and build a normal-

ized term vector $\overrightarrow{t_c} = dim[n]$ for each interaction category c, where $dim[i] = \frac{freq_{i,c}}{\sum_i freq_{i,c}} (1 \le i \le n)$ and $freq_{i,c}$ indicates the frequency of the i-th term in the snippets of interaction category c. Finally, for each new interaction record I, we take advantage of search engines for building its normalized term vector $\overrightarrow{t_I}$ in a similar way and calculate the cosine distances between $\overrightarrow{t_I}$ and all $\overrightarrow{t_c}$, and then I will be transformed to c^* where $c^* = \arg \min_c Distance(\overrightarrow{t_I}, \overrightarrow{t_c})$.

4. EXTRACTING HYPER BEHAVIOR PATTERNS

Though we can reduce the sparseness of behavior patterns by normalizing raw context logs, the total number of behavior patterns for all users is still usually too big compared with the average number of behavior patterns for individual users. For example, in our experimental results, the total number of behavior patterns for all users is 45,230 while the average number of behavior patterns for each user is 308. As a result, the behavior pattern space is still too sparse to be used for comparing different users' habits.

As mentioned in Section 2.2, we may be able to extract some latent and common habits among users from behavior patterns and exploit them for further addressing the sparseness of behavior pattern space. Along this line, we propose to extract mixtures of behavior patterns, which is referred to as *hyper behavior patterns*, for representing these common habits and then using vectors of hyper behavior patterns for representing users' habits. In other words, we try to map users from the behavior pattern space into the hyper behavior pattern space which is much more dense. If we use a $P \times N$ matrix $X = \{x_{ij}\}$ to represent users in the behavior pattern space where x_{ij} indicates the confidence of behavior pattern p_i for user u_j, N and P indicate the total numbers of users and behavior patterns, respectively, the problem can be formally defined as the following Matrix Factorization problem:

$$X = WH + \epsilon,$$

where $W = \{w_{ik}\}$ denotes a $P \times D$ matrix which represents D hyper behavior patterns by D mixtures of behavior patterns, $H = \{h_{kj}\}$ denotes a $D \times N$ matrix which represents N users' habits by N mixtures of hyper behavior patterns, and ϵ denotes the residual noise. The task is finding proper W, H, and ϵ given X. Moreover, we still have three additional constraints for possible solutions as follows:

> **Constraint 1:** All elements in W should be in $[0, 1]$. It is because each column in W represents a hyper behavior pattern by a mixture of original behavior patterns and the element w_{ik} indicates the correlation index between the k-th hyper behavior pattern and the i-th behavior pattern.

> **Constraint 2:** All elements in H should be in $[0, 1]$. It is because each column in H represents a user's habit by a mixture of hyper behavior patterns and the element h_{kj} indicates the correlation index between the k-th hyper behavior pattern and the j-th user.

> **Constraint 3:** Finally, we require $\forall_{j \in [1,N]} \sum_{k=1}^{D} h_{kj} = 1$ for obtaining normalized vectors of hyper behavior patterns for users, which eases the following similarity calculation.

According to above constraints, we use a constraint-based Bayesian Matrix Factorization model [17] which can perform matrix factorization with multiple inequality constraints and equality constraints. In Bayesian Matrix Factorization models, the objective is not to find single solutions for W and H. Instead, we aim to use the model for inferring their posterior probabilistic distributions under a set of model assumptions. These assumptions are specified by the likelihood function $p(X|W, H, \epsilon)$, which expresses the probability of the observed data matrix X given the parameters of the model, i.e., the factorizing matrices W, H and the noise variance ϵ, and the prior $p(W, H)$ and $p(\epsilon)$, which describe available knowledge of the model before observing the data matrix.

In order to keep the formulation general while allowing efficient inference based on Gibbs sampling, we choose an i.i.d. zero mean Gaussian noise model as follows:

$$p(\epsilon_{ij}) = N(\epsilon_{ij}|0, \nu_{ij}) = \frac{1}{\sqrt{2\pi\nu_{ij}}} exp(-\frac{\epsilon_{ij}^2}{2\nu_{ij}}), \qquad (1)$$

where we assume that there is just one overall variance for each matrix element. This noise model leads to the conditional probability of X given the parameters of the model as follows:

$$p(X|W, H, \{\nu_{ij}\}) = \prod_{i=1}^{P}\prod_{j=1}^{N} N(x_{ij}|\sum_{k=1}^{D} w_{ik}h_{kj}, \nu_{ij})$$

$$= \prod_{i=1}^{P}\prod_{j=1}^{N} \frac{1}{\sqrt{2\pi\nu_{ij}}} exp(-\frac{(x_{ij} - \sum_{k=1}^{D} w_{ik}h_{kj})^2}{2\nu_{ij}}) \qquad (2)$$

For the variance parameters, we choose conjugate inverse-gamma priors,

$$p(\nu_{ij}) = IG(\nu_{ij}|\alpha, \beta) = \frac{\beta^{\alpha}}{\Gamma(\alpha)} \nu_{ij}^{-(\alpha+1)} exp(\frac{-\beta}{\nu_{ij}}) \qquad (3)$$

Before defining the prior distributions for the factorizing matrices W and H, we specify the matrices by vectors $\mathbf{w} = vec(W^T) = [w_{11}, w_{12}, \ldots, w_{PD}]^T$ and $\mathbf{h} = vec(H) = [h_{11}, h_{21}, \ldots, h_{DN}]^T$, a Gaussian prior over $\mathbf{w}$ and $\mathbf{h}$ is chosen subject to inequality constraints $\mathbf{Q}$ and equality constraints $\mathbf{R}$ as follows:

$p(\mathbf{w}, \mathbf{h}) \propto$

$$\begin{cases} N\left(\begin{bmatrix} \mathbf{w} \\ \mathbf{h} \end{bmatrix} | \underbrace{\begin{bmatrix} \mu_w \\ \mu_h \end{bmatrix}}_{\equiv\mu}, \underbrace{\begin{bmatrix} \Sigma_w & \Sigma_{wh} \\ \Sigma_{wh}^T & \Sigma_h \end{bmatrix}}_{\equiv\Sigma}\right), \text{if } \mathbf{Q}(\mathbf{w}, \mathbf{h}) \le 0, \mathbf{R}(\mathbf{w}, \mathbf{h}) = 0, \\ 0, \qquad\qquad\qquad\qquad\qquad \text{otherwise}, \end{cases}$$
$$(4)$$

where the constraints $\mathbf{Q} : \mathbb{R}^{PD} \times \mathbb{R}^{DN} \longrightarrow \mathbb{R}^{N_Q}$, $\mathbf{R} : \mathbb{R}^{PD} \times \mathbb{R}^{DN} \longrightarrow \mathbb{R}^{N_R}$ are biaffine maps, which define N_Q inequality constraints, i.e., $\forall_{i\in[1,P],k\in[1,D]} 1 \ge w_{ik} \ge 0$, $\forall_{k\in[1,D],j\in[1,N]} h_{kj} \ge 0$, and N_R equality constraints, i.e., $\forall_{j\in[1,N]} \sum_{k=1}^{D} h_{kj} = 1$. Obviously, we have $N_Q = 2PD + DN$ and $N_R = N$.

The m-th$(1 \le m \le N_Q)$ inequality constraint can be expressed in the form as follows:

$$\mathbf{Q}_m(\mathbf{w}, \mathbf{h}) = q_m + \mathbf{w}^T\mathbf{q}_m^{(w)} + \mathbf{h}^T\mathbf{q}_m^{(h)} + \mathbf{w}^T\mathbf{Q}_m^{(wh)}\mathbf{h} \le 0, \quad (5)$$

For example, the inequality constraint $w_{11} \ge 0$, can be expressed as $\mathbf{Q}_1(\mathbf{w}, \mathbf{h}) = 0 + \mathbf{w}^T[-1, 0, \ldots, 0]_{PD\times1}^T + \mathbf{h}^T[0, 0, \ldots, 0]_{DN\times1}^T + \mathbf{w}^T[\mathbf{0}, \mathbf{0}, \ldots, \mathbf{0}]_{PD\times DN}\mathbf{h} \le 0$

Along this line, the N_Q inequality constraints can be defined in matrix form by rearranging terms as follows:

$$\underbrace{[\mathbf{q}_1^{(w)} + \mathbf{Q}_1^{(wh)}\mathbf{h}, \ldots, \mathbf{q}_{N_Q}^{(w)} + \mathbf{Q}_{N_Q}^{(wh)}\mathbf{h}]^T}_{\equiv\mathbf{Q}_w} \mathbf{w} \le \underbrace{\begin{bmatrix} -q_1 - \mathbf{h}^T\mathbf{q}_1^{(h)} \\ \vdots \\ -q_{N_Q} - \mathbf{h}^T\mathbf{q}_{N_Q}^{(h)} \end{bmatrix}}_{\equiv\mathbf{q}_w}$$
$$(6)$$

Therefore, the inequality constraints $\mathbf{Q}$ can be expressed as $\mathbf{Q}_w^T\mathbf{w} \le \mathbf{q}_w$. Similarly, the equality constraints $\mathbf{R}$ can be expressed as $\mathbf{R}_w^T\mathbf{w} = \mathbf{r}_w$ as well.

To get the factorizing matrices, W and H, we need to infer their posterior distributions. After specifying the model and priors, we can write the posterior distribution of the parameters conditioned on the observed data and hyper-parameters as follows:

$$p(\mathbf{w}, \mathbf{h}, \{\nu_{ij}\}|X, \alpha, \beta, \mu, \Sigma, \mathbf{Q}, \mathbf{R})$$

$$\propto p(\mathbf{w}, \mathbf{h})p(X|\mathbf{w}, \mathbf{h}, \{\nu_{ij}\})\prod_{i=1}^{P}\prod_{j=1}^{N} p(\nu_{ij}) \qquad (7)$$

In practice, it is usually infeasible to obtain the exact solution of the above formula due to the big computation cost. Thus, we use an efficient Markov chain Monte Carlo inference procedure based on Gibbs sampling [17] to approximate the posterior distributions in the following way. Considering the noise variances ν_{ij}, the posterior density is an inverse-gamma distribution due to the choice of conjugate prior

$$p(\nu_{ij}|X, \mathbf{w}, \mathbf{h}) = IG(\nu_{ij}|\overline{\alpha}, \overline{\beta}), \qquad (8)$$

$$\overline{\alpha} = \alpha + \frac{1}{2}, \overline{\beta} = \beta + \frac{1}{2}(x_{ij} - \sum_{k=1}^{D} w_{ik}h_{kj})^2, \qquad (9)$$

from which samples can be generated using standard acceptance-rejection methods.

Since the sampling procedure for $\mathbf{h}$ is identical to that for $\mathbf{w}$ due to the symmetry of the model, we only give the expression for the posterior conditional density of $\mathbf{w}$ as follows:

$$p(\mathbf{w}|X, \mathbf{h}, \{\nu_{ij}\}) \propto \begin{cases} N(\mathbf{w}|\overline{\mu}_w, \overline{\Sigma}_w), \text{if } \mathbf{Q}_w^T\mathbf{w} \le \mathbf{q}_w, \mathbf{R}_w^T\mathbf{w} = \mathbf{r}_w, \\ 0, \qquad\qquad \text{otherwise} \end{cases}$$
$$(10)$$

$$\overline{\Sigma}_w = (\widetilde{\Sigma}_w^{-1} + \widetilde{\mathbf{H}}\mathbf{V}^{-1}\widetilde{\mathbf{H}}^T)^{-1}, \overline{\mu}_w = \overline{\Sigma}_w(\widetilde{\Sigma}_w^{-1}\widetilde{\mu}_w + \widetilde{\mathbf{H}}\mathbf{V}^{-1}\mathbf{x}),$$
$$(11)$$

where $\widetilde{\mu}_w = \mu_w + \Sigma_{wh}\Sigma_h^{-1}(\mathbf{h} - \mu_h)$, $\widetilde{\Sigma}_w = \Sigma_w - \Sigma_{wh}\Sigma_h^{-1}\Sigma_{wh}^T$, $\mathbf{V} = diag(\nu_{11}, \nu_{12}, \ldots, \nu_{PN})$, $\widetilde{\mathbf{H}} = diag(H, \ldots, H)$ is a diagonal block matrix with P repetitions of H, and $\mathbf{x} = vec(X^T) = [x_{11}, x_{12}, \ldots, x_{PN}]^T$

Finally, we can perform Gibbs sampling iteratively as follows:

a. the noise variance is generated from the inverse-gamma density in Eq. 8;

b. $\mathbf{w}$ is generated from the constrained Gaussian density in Eq. 10;

c. $\mathbf{h}$ is generated from a constrained Gaussian analogous to Eq. 10.

Another important problem for extracting hyper behavior patterns is the selection of the hyper behavior pattern

number, i.e.,D, for a given observed X. In the Bayesian framework, model selection can be performed by evaluating the marginal likelihood, $p(X)$, which involves an intractable integral over the posterior and thus is usually estimated but not calculated exactly. In this paper, we use Chib's method [6] for estimating $p(X)$, which can be implemented directly using the described Gibbs sampler.

5. EXPERIMENTS

In this section, we conduct extensive experiments for e-valuating the effectiveness of our approach for reducing the sparseness of behavior pattern space and discovering similar mobile users with respect to their habits.

5.1 Data Set

The experimental data set is collected from many volunteers by a major manufacturer of smart mobile devices. The data set contains rich context data including time information, profile information and cell ID trajectories, and interaction records of 443 mobile users. The data set contains 8,852,187 context records where 1,097,189 of them contain non-empty interaction records. Figures 4 illustrates the distribution of time lengths for each user's data.
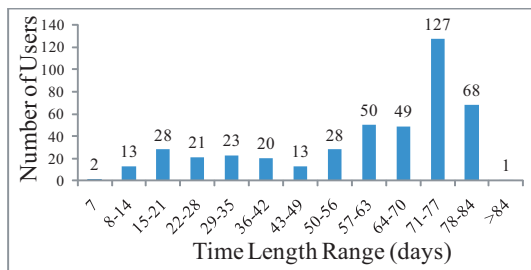

Figure 4: The time length distribution of the collected data for all volunteers.

5.2 Preparation

In order to normalize the raw context log of each user, we first mine their homes and work places from their cell ID trajectories and remove the cell IDs which cannot be mapped to either "Home" or "Work Place". The maximum number of unique cell IDs in each closure pattern is empirically set to 2 according to [23]. The time ratio thresholds of "Home" and "Work Place" are empirically set to be 0.3 and 0.5, respectively. Second, we map raw interaction records into interaction categories by defining 13 interaction categories based on the taxonomy of Nokia Ovi Store (www.ovi.com) and then build term vectors for each interaction category. For each interaction category, we manually select five typical interaction records from the collected data as category seeds and then leverage a search engine to enrich its textual description. The 13 predefined interaction categories and the corresponding category seeds as shown in Table 2.

We take advantage of the GCPM algorithm [3] to mine behavior patterns from the normalized context logs. For each context log, the min_sup is set according to the spanning time to ensure that each pattern is followed by the user at least averagely once a week. For example, if a context log spans for three weeks, the min_sup is set to be three. The min_conf is uniformly set to be 0.6.

Table 2: The interaction categories predefined in our experiments.

Interaction Categories	Category Seeds
Message	Msg. reader, AgileMessenger, ShakeSMS Animated SMS, Audio message
Call	FlipSilent, MobiGenie, TastePhone Telephone, TiViPhone
Mail	ProfiMail, Z7EmailViewer, Email Google Mail, Gmail
Web Service	live!, Live TV, Symella SmartGuard, web
Multimedia	DivX Player, SPODTRONIC, kyte mobbler, CorePlayer
Utility	KAV Mobile, CarUIApp, Kalendarz Mobile Security Update, cClock
Navigation	CellPos, Mapas, GlastoNav Map24 Mobile, psdk_lmsrvapp
System	Theme DIY, Mass memory, SmartProfile MemoryUp, Device status
Reference	ViewRanger, BibleReader, MyNokia Mobile Weather, ForecaWeather
SNS	IM+, Lifeblog, SmsChat Online share, EQO Mobile
Game	Solytare, JewelQuest2, JbakTaskMan SuperMario, Guitar Hero III,
Management	File manager, PhoNetInfo, MessageList Handy Blacklist, Birthday Manager
Business	Truphone, SlideHelper, Team suite Spreadbet Marketmaker, Adobe Reader

Table 3: A example of hyper behavior pattern.

{(Day period: Morning),(Time range: AM8:00-9:00)} ⟶ Multimedia
{(Is a holiday?: No),(Time range: AM8:00-9:00)}⟶ Multimedia
{(Day period: Morning),(Time range: AM8:00-9:00), (Profile: General)}⟶ Multimedia
{(Profile: General)} ⟶ Multimedia
{(Is a holiday?: No),(Day period: Morning), (Time range: AM8:00-9:00)} ⟶ Multimedia
{(Day name: Thursday),(Is a holiday?: No),(Profile: General)} ⟶ Multimedia
{(Day name: Friday),(Is a holiday?: No),(Profile: General)} ⟶ Multimedia
{(Is a holiday?: No),(Day period: Morning), (Time range: AM8:00-9:00),(Profile: General)}⟶ Multimedia
{(Is a holiday?: Yes),(Day period: Morning),(Profile: General)} ⟶ Multimedia
{(Is a holiday?: No),(Time range: AM8:00-9:00), (Profile: General)}⟶ Multimedia

Given the mined behavior patterns, we build the user-behavior pattern matrix and perform constraint-based Matrix Factorization on it. The number of hyper behavior patterns is set to be 15 empirically according to the method mentioned in Section 4. Table 3 shows an example of hyper behavior pattern mined from the normalized context logs which indicates a common habit that "users usually use multimedia applications between AM8:00-9:00 in work day mornings". Limited by space, we just list the top 10 behavior patterns in the hyper behavior pattern.

5.3 Dealing with the Sparseness of Behavior Patterns

To evaluate the effectiveness of normalizing context logs for reducing the sparseness of behavior pattern space, we first compare the sparseness of behavior pattern spaces for raw context logs and normalized context logs. Figure 5 shows the distributions of behavior patterns with respect to the number of their following users for raw context logs and normalized context logs, respectively. To observe the differ-

ence of the distributions more clearly, we mark the average number of users (Avg. Num. of Follow. Users) who follow one same pattern, which is calculated by $\frac{\sum_{i=1}^{N_p} n_i}{N_p}$, where N_p indicates the total number of unique behavior patterns, and n_i indicates the number of users who follow the i-th pattern. From these figures, we can see that the space of the behavior patterns mined from normalized context logs is much more dense than that of the behavior patterns mined from raw context logs. To be specific, in the space of behavior patterns mined from raw context logs, the average number of users who follow one same pattern is 1.13 and at most 13 users follow one same pattern. In contrast, after normalizing context logs, the average number of users who follow one same pattern becomes 3 and there are at most 47 users who follow one same pattern.
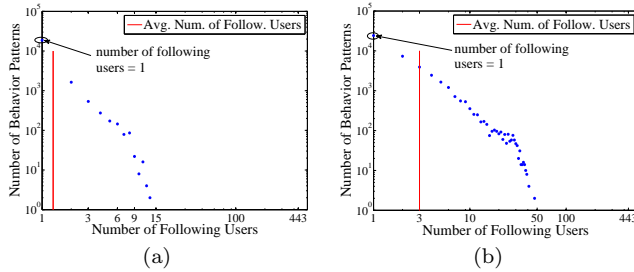

Figure 5: The distributions of behavior patterns with respect to the number of their following users for (a) raw context logs and (b) normalized context logs.

However, as mentioned in Section 2.2, the space of behavior patterns mined from normalized context logs is still too sparse to be directly used for discovering similar users with respect to their habits, we further take advantage of the Matrix Factorization technology for extracting hyper behavior patterns which are in an extremely low dimension space. One may argue for an alternative approach which directly perform Matrix Factorization on the behavior patterns mined from raw context logs. Although this approach can also significantly deal with the sparseness problem, some important information about the similarity of user habits will be missed due to the too specific raw context data and therefore the performance for discovering similar users with respect to their habits may be affected. Actually, the experimental results in the following section demonstrate the concern. Therefore, normalizing context logs is useful not only in terms of reducing the sparseness of behavior pattern space, but also in terms of capturing more information about the similarity with respect to user habits.

5.4 Effectiveness of Discovering Similar Users with Respect to Their Habits

To evaluate the effectiveness of our approach for discovering similar users with respect to their habits, we randomly select several groups of test users and each group contains several users. For each test user, our approach and baselines find the top K similar users with respect to their habits. The results are evaluated manually because there is no ground truth for the experimental data. Due to the expensiveness of manual evaluation, the scale of the experiments is limited. Specially, we randomly select two user groups where each

group contains 10 users for our approach and three baselines while the maximum K is set to be 5, which means we label 400 pairs of users by considering their behavior patterns mined from their normalized context logs. The details of the selected baselines, metrics and experimental results are presented in the following sections.

5.4.1 Baselines

Since the problem of leveraging behavior patterns for discovering similar users with respect to their habits is under-explored, it is difficult to directly find existing works as baselines. Alternatively, we extend Cao et al. [3]'s work as one baseline. Moreover, in this study, to address the sparseness of behavior pattern space, our approach includes two major stages, namely, normalizing context logs and extracting hyper behavior patterns. To separately evaluate the contribution of each stage for discovering similar users with respect to their habits, we design two other baselines which only take advantage of one stage for improving the ability of discovering similar users with respect to their habits. Finally, the baselines used in our experiments are as follows.

R-BP: Mine behavior patterns from raw context logs and then calculate the similarity with respect to user habits by the cosine distance between the behavior pattern vectors of users. It is extended from [3].
N-BP: Mine behavior patterns from normalized context logs and then calculate the similarity of user habits by the cosine distance between the behavior pattern vectors of users.
R-BP+CMF: Mine behavior patterns from raw context logs and extract hyper behavior patterns through constraint-based Matrix Factorization, then calculate the similarity of user habits by the cosine distance between the hyper behavior pattern vectors of users.

For simplicity, we denote our approach as **N-BP+CMF** which stands for mining **B**ehavior **P**atterns from **N**ormalized context logs and extract hyper behavior patterns through **C**onstraint based **M**atrix **F**actorization.

5.4.2 Metrics

Since there is no ground truth for evaluation, for each test user, we invite 10 college volunteers to manually evaluate the experimental results of similar users with respect to their habits. All of them are graduate students major in Computer Science. The judges are provided the behavior patterns of both the given user and candidate similar users to regard the real similarity from their perspective. These behavior patterns are mined from normalized context logs and organized by interaction categories for clarity. For each interaction category, the top 20 behavior patterns with highest confidence are selected. For each test result, the judgers can give a score from zero to two, where zero means completely dissimilar, one means some habits are similar and two indicates a high similarity for all habits. To avoid bias, the test results from different approaches are mixed and each test result is judged by at least three judgers.

The problem of discovering similar users with respect to their habits can be regarded as ranking users in order of their habit similarity. Therefore, we can borrow some classical metrics used in the field of information retrieval though the latter focuses on ranking relevant documents but not similar

users. Alone this line, we use the following metrics in our experiments.

Precision at K **(P@K)** indicates the ratio of the returned users that are labeled as partly similar or very similar in the top K results of a ranked list.

Normalized Discounted Cumulative Gain(nDCG) is devised specifically for web search evaluation[10]. For a given user, the ranked K candidate similar users are examined from the top to down. nDCG@K is computed by $nDCG@K = \frac{DCG@K}{IDCG@K}$, where $DCG@K = \sum_{j=1}^{K} (2^{r(j)}-1)/log_2 (1+j)$, $r(j)$ indicates a similarity score from zero to two, and IDCG@K is the DCG@K value of ideal ranking list of the returned result.

Mean Average Precision (MAP) is computed by $\frac{1}{|N|} \sum_{i=1}^{N} \frac{\sum_{r=1}^{K}(P_i(r) \times rel_i(r))}{|R_i|}$, where N indicates the number of test users and $|R_i|$ indicates the number of found similar users for the test user U_i. r denotes a given cut-off rank, $P_i(r)$ denotes the precision of U_i at a given cut-off rank r where both partly similar users and totally similar users are regarded correct, and $rel_i(r)$ is the binary function on the correctness of a given rank r.

5.4.3 Overall Experimental Results

The comparison results between our approach and the baselines for the two groups of test users in terms of Precision, nDCG and MAP are reported in Figure 6, Figure 7, and Figure 8, respectively. From these figures, we can see that N-BP+CMF consistently outperforms R-BP+CMF while N-BP consistently outperforms R-BP in terms of all metrics, which implies that normalizing raw context logs can improve the performance of behavior pattern based similar user discovery. Moreover, we can also see that N-BP+CMF consistently outperforms N-BP in terms of all metrics, which implies that extracting hyper behavior patterns can improve the performance of behavior pattern based similar user discovery as well.
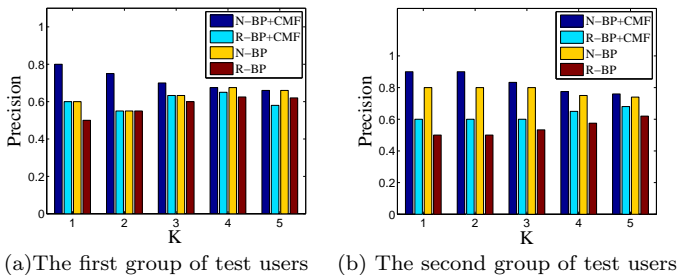

(a)The first group of test users   (b) The second group of test users

Figure 6: Precision at varying K**.**

5.5 Effectiveness of Segmenting Users with Respect to Their Habits

Except for directly evaluating the proposed approach for discovering similar users, we also study its effectiveness of segmenting users with respect to their habits. To be specific, we cluster the users according to their similarities which are calculated by the cosine distances between their hyper behavior pattern vectors and behavior patterns, respectively, and compare their performance. Both the hyper behavior patterns and behavior patterns are extracted from
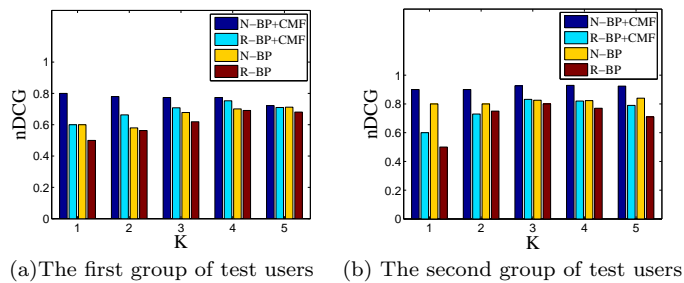

(a)The first group of test users   (b) The second group of test users

Figure 7: nDCG at varying K**.**

(a)The first group of test users   (b) The second group of test users
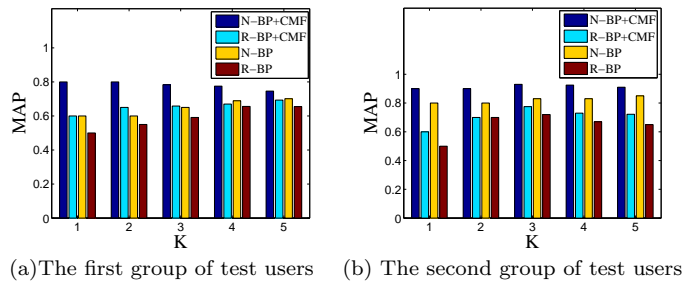
Figure 8: MAP at varying K**.**

normalized context logs. We utilize a clustering algorithm proposed in [5] which does not require a parameter to indicate the number of clusters but only needs a parameter to indicate the minimum average mutual similarity S_{min} for the data points in each cluster. The average mutual similarity for a user cluster C is calculated as $S_C = \frac{2 \times \sum_{1 \le i < j \le N_C} Sim(u_i,u_j)}{N_C \times (N_C-1)}$, where N_C indicates the number of users in C and $Sim(u_i,u_j)$ denotes the similarity between the i-th user and the j-th user in C.

For the clusters based hyper behavior pattern vectors, S_{min} is empirically set to be 0.9. However, for the clusters based on behavior patterns, it is difficult to select a proper S_{min} because there exist rare pairs of users whose similarities are relatively big when the similarity is calculated based on the sparse behavior pattern space. Through several trials, we determine $S_{min} = 0.3$ for those clusters because in this case the results look relatively good.

In order to intuitively compare the segmentation results, we exploit Principal Component Analysis (PCA) to map the 443 users represented by high dimensional behavior pattern vectors to the space of top two principal components. Figure 9(a) and (b) show the segmentation results of 443 mobile users by utilizing hyper behavior patterns and behavior patterns, respectively. In these figures each color denotes a cluster of users with respect to their habits and the outliers are removed for the clarity. From the two figures we can see that the segmentation results based on hyper behavior patterns look better because the clusters are more balanced in size. Moreover, the segmentation results based on hyper behavior patterns only have 14 outliers while the other results have 84 outliers, which further implies that hyper behavior patterns are more effective for measuring the similarity between users with respect to their habits.

We also study each cluster of users based on hyper behavior patterns and find all of them have obvious relevance to particular hyper behavior patterns. For example, Figure 10 (a) and (b) show the values of each hyper behavior pattern dimension for the users in two randomly selected clusters
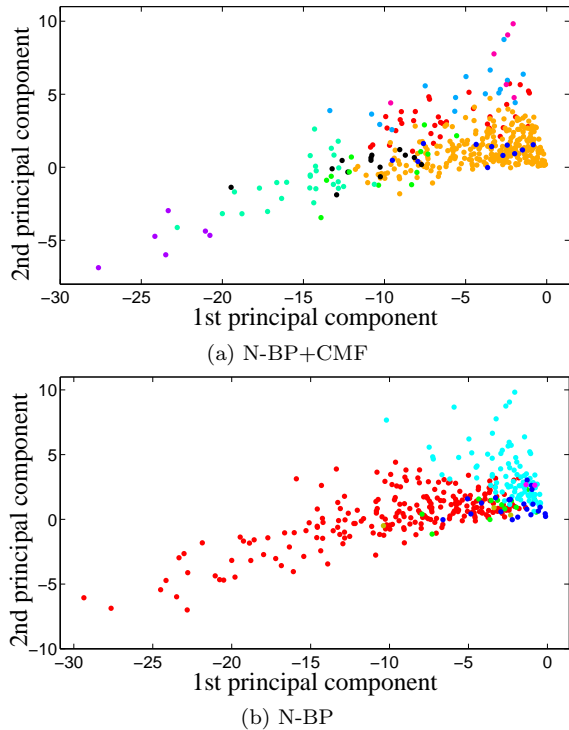
(a) N-BP+CMF

(b) N-BP

Figure 9: The segmentation results of mobile users by utilizing (a) hyper behavior patterns (N-BP+CMF) and (b) behavior patterns (N-BP).

C_a and C_b with box plots. From these figures we can clearly see that the users in the two clusters dramatically have high values in particular hyper behavior pattern dimensions, which imply they follow the corresponding hyper behavior patterns. To be specific, the users in C_a seem to follow the 14-th hyper behavior pattern which indicates a common habit that "the users usually call others in afternoons while the profile is set to be General", and the users in C_b seem to follow the 11-th hyper behavior pattern which indicates a common habit that "the users usually send emails on work days while the profile is set to be Meeting".
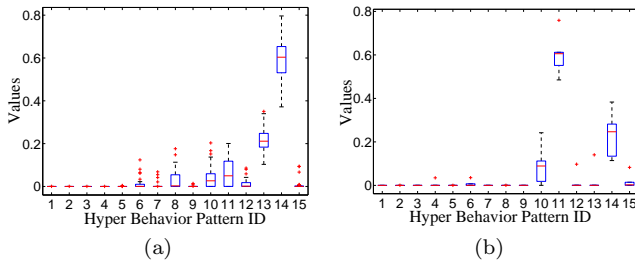

(a)

(b)

Figure 10: The values of each hyper behavior pattern dimension for the users in (a) C_a and (b) C_b with box plots

6. RELATED WORK

The main idea of habit based user similarity measurement is to calculate the user similarity by mining and comparing user habits. Along this line, many previous studies take into account user profiles or online user behaviors in Web communities for understanding user habits. For example,

Hopfgartner et al.[9] exploited implicit user interaction to capture long-term user habits in a profile that was used to retrieve and recommend news stories to the users. Mobasher et al. [16] presented two techniques, based on clustering of user transactions and clustering of page views, to discover overlapping aggregate profiles for describing user habits. Zhou et al. [27] and Wu et al. [21] modeled user habits as latent variables and extracted them from online behaviors through the probability mixture model. However, as pointed out by [24], the interests in the virtual world may be not capable of capturing the "real" characteristics of user behaviors.

With the increasing popularity of smart mobile devices and their more and more powerful ability of sensing user contexts, richer contextual information and activity information of users can be logged for mining user habits, which provides a more effective way of capturing the "real" characteristics of user behaviors. Previous studies on mining mobile user habits can be roughly classified into two categories as follows.

The first category only leverages location-based context and time information. In some of these works, user similarity is defined based on the geographic features of mobile users' trajectories. For example, in [14] and [26], both the sequence property of people's movement behaviors and the hierarchy property of geography spaces were taken into account for consistently modeling each individual's location and effectively measuring the similarity among users. Lu et al. [15] proposed a LBS-Alignment method to calculate the similarity of two mobile users by using the longest common sequence within their mobile sequential patterns. In other works, the semantic information of locations is considered in measuring the similarity between different users. The common point of these works is modeling users' trajectories with semantic location histories, e.g., shopping malls → restaurants → cinemas. However, they are different in defining the similarity measurement of semantic trajectories. For example, Xiao et al.[22] presented a maximal travel math (MTM) algorithm for measuring the similarity of semantic location history. Ying et al.[24] proposed Maximal Semantic Trajectory Pattern Similarity(MSTP-Similarity) to measure the semantic similarity between trajectories.

The second category includes some recent works which consider rich contextual information including user interactions. For example, Tseng et al. [18, 19, 20] studied the problem of discovering user habits by considering both location information and service requests in mobile web environments. However, these works do not take into account other context information such as transportation status and phone profile, which are useful for mining user habits as well. Karatzoglou et al.[11] proposed multiverse recommendation based on the Tucker tensor factorization model that allows for a flexible and generic integration of contextual information. However, this model is not scalable well since its complexity is exponential for the number of context variables and polynomial for the size of factorizations. Additionally, it only works for categorical context variables and is sensitive to noisy data. To effectively exploit the rich contextual information and user interactions, Cao et al.[3] proposed to mine associations between interactions and contexts, i.e., behavior patterns, from context logs of mobile users for representing user habits. However, leveraging mined behavior patterns for discovering similar users remains under-exploited due

to the extreme sparseness of behavior pattern space, which motivates our work represented in this paper.

7. CONCLUSION AND FUTURE WORK

In this paper, we proposed a two-stage approach for addressing the sparseness of behavior pattern space and thus made possible to leverage behavior patterns for discover similar mobile users with respect to their habits. Specifically, we first normalized the raw context logs by transforming location-based context data to social locations and transforming raw interaction records to interaction categories. Secondly, we took advantage of a constraint-based Matrix Factorization model for extracting hyper behavior patterns from the behavior patterns mined from normalized context logs and represented user habits by vectors of hyper behavior patterns. The experiments conducted on real data sets show that our approach outperforms three baselines in terms of the effectiveness of discovering similar mobile users with respect to their habits.

As mentioned above, discovering similar users with respect to their habits plays an important role in a wide range of real applications. In this paper, we preliminarily investigate the performance of the proposed approach for user segmentation. As for future work, we plan to study how well the novel approach will improve more relevant applications such as collaborative filtering for recommendation.

8. ACKNOWLEDGMENTS

9. REFERENCES

[1] R. Agrawal and R. Srikant. Fast algorithms for mining association rules. In *Proceedings of the 20th International Conference on Very Large Data Bases (VLDB'94)*, pages 487–499, 1994.

[2] K. S. Beyer, J. Goldstein, R. Ramakrishnan, and U. Shaft. When is "nearest neighbor" meaningful? In *Proceedings of the 7th International Conference on Database Theory*, ICDT'99, pages 217–235, 1999.

[3] H. Cao, T. Bao, Q. Yang, E. Chen, and J. Tian. An effective approach for mining mobile user habits. In *Proceedings of the 19th ACM Conference on Information and Knowledge Management (CIKM'10)*, pages 1677–1680, 2010.

[4] H. Cao, D. H. Hu, D. Shen, D. Jiang, J. Sun, E. Chen, and Q. Yang. Context-aware query classification. In *Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval*, SIGIR '09, pages 3–10, 2009.

[5] H. Cao, D. Jiang, J. Pei, Q. He, Z. Liao, E. Chen, and H. Li. Context-aware query suggestion by mining click-through and session data. In *KDD*, pages 875–883, 2008.

[6] S. Chib. Marginal likelihood from the gibbs output. *Journal of the American Statistical Association*, 90(432):1313–1321, 1995.

[7] A. Das, M. Datar, A. Garg, and S. Rajaram. Google news personalization: scalable online collaborative filtering. In *WWW*, pages 271–280, 2007.

[8] J. Han, J. Pei, Y. Yin, and R. Mao. Mining frequent patterns without candidate generation: A frequent-pattern tree approach. *Data Min. Knowl. Discov.*, 8(1):53–87, 2004.

[9] F. Hopfgartner and J. M. Jose. Semantic user profiling techniques for personalised multimedia recommendation. *Multimedia Syst.*, 16(4-5):255–274, 2010.

[10] K. Järvelin and J. Kekäläinen. Ir evaluation methods for retrieving highly relevant documents. In *SIGIR*, pages 41–48, 2000.

[11] A. Karatzoglou, X. Amatriain, L. Baltrunas, and N. Oliver. Multiverse recommendation: n-dimensional tensor factorization for context-aware collaborative filtering. In *RecSys*, pages 79–86, 2010.

[12] C. Kirchmeyer. Demographic similarity to the work group: A longitudinal study of managers at the early career stage. *Journal of Organizational Behavior*, 16(1):67–83, 1995.

[13] K. Laasonen, M. Raento, and H. Toivonen. Adaptive on-device location recognition. In *Pervasive*, volume 3001, pages 287–304, 2004.

[14] Q. Li, Y. Zheng, X. Xie, Y. Chen, W. Liu, and W.-Y. Ma. Mining user similarity based on location history. In *GIS*, page 34, 2008.

[15] E. H.-C. Lu and V. S. Tseng. Mining cluster-based mobile sequential patterns in location-based service environments. In *Mobile Data Management*, pages 273–278, 2009.

[16] B. Mobasher, H. Dai, T. Luo, and M. Nakagawa. Discovery and evaluation of aggregate usage profiles for web personalization. *Data Mining and Knowledge Discovery*, 6:61–82, 2002.

[17] M. Schmidt. Linearly constrained bayesian matrix factorization for blind source separation. In Y. Bengio, D. Schuurmans, J. Lafferty, C. K. I. Williams, and A. Culotta, editors, *Advances in Neural Information Processing Systems 22*, pages 1624–1632. 2009.

[18] S.-M. Tseng and C.-F. Tsui. An efficient method for mining associated service patterns in mobile web environments. In *SAC*, pages 455–459, 2003.

[19] S.-M. Tseng and C.-F. Tsui. Mining multilevel and location-aware service patterns in mobile web environments. *IEEE Transactions on Systems, Man, and Cybernetics, Part B*, 34(6):2480–2485, 2004.

[20] V. S. Tseng and K. W. Lin. Efficient mining and prediction of user behavior patterns in mobile web systems. *Information and Software Technology*, 48(6):357–369, 2006.

[21] X. Wu, J. Yan, N. Liu, S. Yan, Y. Chen, and Z. Chen. Probabilistic latent semantic user segmentation for behavioral targeted advertising. In *KDD Workshop on Data Mining and Audience Intelligence for Advertising*, pages 10–17, 2009.

[22] X. Xiao, Y. Zheng, Q. Luo, and X. Xie. Finding similar users using category-based location history. In *GIS*, pages 442–445, 2010.

[23] G. Yang. Discovering significant places from mobile phones - a mass market solution. In *MELT*, pages 34–49, 2009.

[24] J. J.-C. Ying, E. H.-C. Lu, W.-C. Lee, T.-C. Weng, and V. S. Tseng. Mining user similarity from semantic trajectories. In *GIS-LBSN*, pages 19–26, 2010.

[25] V. W. Zheng, B. Cao, Y. Zheng, X. Xie, and Q. Yang. Collaborative filtering meets mobile recommendation: A user-centered approach. In *AAAI*, 2010.

[26] Y. Zheng, L. Zhang, Z. Ma, X. Xie, and W.-Y. Ma. Recommending friends and locations based on individual location history. *TWEB*, 5(1):5, 2011.

[27] K. Y. Zhou and B. Mobasher. Web user segmentation based on a mixture of factor Analyzers. In *Proceedings of the 7th International Conference on E-Commerce and Web Technologies (EC-Web'06)*, pages 11–20, 2006.