Context-aware Collaborative Filtering in Location Based Services

Haosheng Huang

Vienna University of Technology

haosheng.huang@gmail.com

Abstract

Technical advances in mobile devices and mobile communication have led to the introduction of Location Based Services (LBS). Currently, providing context-aware services/information is still a key challenge in LBS applications. Collaborative filtering (CF), known as Amazon-like recommendation, is a promising solution for providing context-aware recommendations. The goal of this report is to investigate how context-aware CF (CaCF) can be introduced into LBS to provide context-aware recommendations. Specifically, we focus on applying CaCF methods on the highly available spatio-temporal trajectories to enhance visitors with context-aware POI (Point of Interest) recommendations.

This report first proposes a two stage method to identify context parameters which are relevant and thus needed to be modeled in CaCF. After identifying relevant context parameters, we explore two different approaches (i.e., Local-Global Approach LGA, and Statistic-Based Approach SBA) to measure similarity between different contexts (situations). In considering two different ways of incorporating context information into the CF process, four CaCF methods are designed for LBS: LGA_CP_CaCF (using LGA and contextual pre-filtering), LGA_CM_CaCF (using LGA and contextual modeling), SBA_CP_CaCF (using SBA and contextual pre-filtering), and SBA_CM_CaCF (using SBA and contextual modelling). With

these CaCF methods, smart services like "in similar context, other people similar to you often …" can be provided.

Finally some experiments are designed to evaluate the proposed methods. The results of the experiments show that the proposed CaCF methods are feasible and useful for providing context-aware recommendation in LBS applications. Also, we prove that including context information in the CF process can improve the predictive performances in LBS.

Key words: context-aware computing, collaborative filtering, context similarity, context-aware recommendation, trajectory

1. Introduction

Recent years have seen raising interest in Location Based Services (LBS) with the continual evolution of mobile devices and communication technology. LBS are becoming ever more accessible not only in city wide outdoor environments but also in shopping malls, museums and other indoor environments. They have been applied in emergency services, tourism services, intelligent transport services, gaming, assistive services, etc (Raper et al. 2007). Among them, mobile guide (mobile tourism service) is the largest group of LBS application. In this report, we mainly focus on mobile guides.

Context-awareness is a key in LBS. Currently, context-awareness in mobile guides mostly relies on an adaptation engine to determine the appropriateness of POIs for satisfying user's needs and context. However, building the adaptation engine has to undergo a long process of knowledge acquisition which is very time-consuming and impractical for lots of LBS.

Additionally, the increasing ubiquity of GPS-enabled devices has led to the collection of large spatio-temporal datasets, such as trajectories. These trajectories may reflect the perspective and experiences of other people who solve their spatial tasks (e.g., choosing which POI to visit next) in this situation. It is obvious in our daily life that experiences from past users (especially similar users) in similar context can help current users to efficiently solve their problems (Wexelblat 1999). Therefore, by aggregating the trajectories, LBS can provide users with smart services, such as providing social affordance for making decisions. However, little research has addressed these considerations.

Collaborative filtering (CF, "Amazon-like recommendation") is a promising solution for the above problems. It uses opinions (i.e., UGC) of similar users in similar context to help the current user efficiently identify information of interest (Resnick and Varian 1997). As a result, by incorporating CF into LBS, relevant information (e.g., POIs in mobile guides) matching user's current situation can be identified (by aggregating opinions from similar users in a similar context). However, little research has addressed these considerations.

The goal of this report is to investigate methods of introducing context-aware CF (CaCF) into LBS to provide context-aware recommendations. Specifically, we aim at applying CaCF methods on the highly available spatio-temporal trajectories to enhance visitors with context-aware POI recommendations. With CaCF, smart services like "in similar context, other people similar to you often visited POI A" can be provided in LBS applications.

The rest of this report is structured as follows. In section 2, we outline related work. Section 3 presents a methodology for incorporating CaCF into LBS, and develops four CaCF methods. The proposed methods are implemented and evaluated with some experiments in section 4. Some major experimental results and related discussions are also provided in this section. Finally, we draw conclusions and present future work in section 5.

2. Related work

2.1 Context-aware Recommendations

There are several surveys on mobile guides in the literature (Baus et al. 2005, Schwinger et al. 2005, Krüger et al. 2007, Raper et al. 2007). According to them, most of the mobile guides provided user-related or context-related adaptive services. For providing context-aware services (e.g., recommending POIs matching the context), current mobile guides mainly relied on knowledge about POIs (domain model, DM), knowledge (preferences and needs) about the user (user model, UM) and her/his context (context model, CM). An adaptation engine measures the appropriateness of the objects (DM) for satisfying a particular user's needs and context (UM and CM), and returns relevant objects.

UM includes information about user's interest, preferences and needs. UM can be static (Baus et al. 2001, Kray 2003) and dynamic (Cheverst et al. 2000, Wiesenhofer et al. 2007). For building DM and the adaptation engine, a long underlying learning (knowledge acquisition) process has to be carried out in the field which is always very time-consuming and impractical for lots of LBS applications. Most importantly, current mobile guides are unable to effectively provide users with context-aware services in an unseen situation or situations with little previous knowledge, which is very common in LBS applications.

In contrast to the above approaches, we aim at using CF to providing context-aware recommendations. In the following, we analyze CF's potential in solving the problems in current context-aware recommendation (i.e., the long process of knowledge acquisition in providing context-awareness).

One of the key goals of context-awareness in mobile guides is to provide a user with relevant information based on her/his context. This goal matches with the goal of CF, especially context-aware CF, which aggregates opinions of similar users in similar context to help individuals efficiently identify interesting information (Resnick and Varian 1997). Opinions from other users reflect their perceptions on the fitness/appropriateness of a particular item (information) for the context. If similar users in similar context like that particular item, this item can be considered as a matching item for the current user in the current context. As a result, CF, especially CaCF, can be viewed as a real-time underlying learning process of building DM and UM, and an automatic engine for identifying relevant information. Moreover, as CF solely relies on user feedback and requires no previous domain knowledge, LBS systems employing CF will be able to effectively provide context-aware services in dynamic environments and unseen situations. Thus, CF can be a novel method of providing context-aware services in LBS.

2.2 Mining Trajectories

With the increasing ubiquity of GPS-enabled devices, more and more people start to record their travel/sports experience with GPS logs, and then upload, visualize and browse their GPS data on a web map. Therefore, large spatio-temporal datasets (e.g., trajectories) are created every day, or even every minute. Recently, mining these kinds of user-generated GPS data is receiving considerable attention.

There are some researches mining personal location history based on individual trajectories. They focus on detecting significant locations of a user, predicting user's behavior among these locations, identifying user's spatio-temporal behavior patterns, and recognizing user's activities on each location (Li et al. 2008). In the meantime, lots of other researches mine multiple users' trajectories to understand mobility-related phenomena, e.g., Gonotti et al. (2007) aggregate a set of many individual trajectories to identify spatio-temporal behavior patterns, Zheng et al. (2008) infer users' transportation mode (e.g., walking and driving) based on GPS trajectories of different users, Zheng et al. (2009) mine interesting locations and travel sequences from multiple users' trajectories. Li et al. (2008) propose an interesting user similarity measure based on different users' trajectories, however, they don't incorporate the similarity measure into CF process.

There are also some researches focusing on using trajectories for recommendations. Takeuchi and Sugimoto (2006) recommend shops to users based on their individual preferences and needs, estimated by analyzing their past location history (i.e., GPS trajectories). Bohnert et al. (2008) develop a system for exhibit recommendation based on users' trajectories in museum. They use two adaptive collaborative models (interest model and transition model), and a combined model for predicting a visitor's next location in a museum. However, it is important to note that context information (except location) is not considered in these researches.

In contrast to the above researches, this paper aims at designing context-aware CF methods to utilize trajectories for context-aware POI recommendations. Incorporating context information into the CF process for LBS applications is the main research focus.

2.3 Context-aware CF (CaCF) in LBS Applications

The process of user-based CF includes three key stages: data collection for building user profiles (generally, in the form of ratings, e.g., a set of preference ratings given by the same user on different POIs), computation of user similarities, and aggregation of ratings from the N most similar users (or users with bigger similarity with the current user) for recommendation. CF is often applied in Web-based applications, such as movie recommendations, and product recommendations (see Amazon.com). Also, context-aware CF in Web-based applications is still very challenging (Adomavicius and Tuzhilin).

There are some tries on applying CF in LBS, such as restaurant recommendation (Horozov et al. 2006), event recommendation (de Spindler et al. 2006, Li et al. 2009), shop recommendation (Takeuchi and Sugimoto 2006), recommendation in museums (Bohnert et al. 2008), and POI recommendation for tourism (van Setten et al. 2004). However, most of the researches only employ location as contextual factor, and don't consider other contextual factors which are also relevant for generating recommendations, e.g., weather and companion (with whom).

Context-awareness is a key when introducing CF into LBS. For example, recommending a place for the same person to visit may vary according to different weather (rainy or sunny) and different companion (alone or with children). "There is more to context than location" (Schmidt et al. 1999). When employing more context information into CF for LBS, the problem becomes very challenging. A comprehensive investigation of how context can be incorporated into LBS-based CF is urgently needed. It is also important to note that none of the research focusing on experimentally studying whether including contextual information in a CF for LBS can improve the recommendation performance.

2.4 Key Challenges of Incorporating Context-aware CF into LBS

Context-aware CF (CaCF) aggregates what similar users chose in similar context in the past for recommendation. Several key issues have to be considered when providing CaCF in LBS: annotating user profiles with context, measuring similarities between contexts, and incorporating context information into the CF process.

Annotating user profiles with context

In CaCF, user profiles should be annotated with context. To be more specific, the context within which the item is rated or visited should be stored along with the rating, i.e., <user, item, rating, context>. Context is used to define a user's preferences (ratings) in different situations. A context can be characterized by a set of context parameters. Not all context parameters are relevant for generating recommendations. In order to annotate user profiles with context, a main question has to be answered: which context parameters are relevant and thus needed to be modeled. Lots of researches choose some features of the world as their context parameters from their own views, e.g., Panniello et al. (2009) employ *"the period of the year"* to recommend products in an e-commerce portal, Adomavicius et al. (2005) use *"time (weekday, weekend)"*, *"company (alone, friend, etc.)"* and *"place (theater, at home)"* to provide context-aware movie recommendations. However, to the best of our knowledge, none of the research in CF explores methods of identifying relevant context parameters.

Measuring context similarity

The goal of measuring context similarity is to determine which ratings (or user profiles, i.e., trajectories in this report) are more relevant for the current context (i.e., more useful for making recommendations for the current context). Chen (2005) proposes to learn context similarity based on the data. The basic assumption is that if the ratings for an item are similar for two different contexts, then these two contexts can be considered as similar. However, she doesn't give any details about the implement and evaluation. In contrast, Panniello et al. (2009) utilize the hierarchical characteristic of context to calculate the context similarity.

However, they only consider a single context parameter (“the period of the year”). There are some related methods on similarity measure in other fields, e.g., local-global principle and machine learning approach by Stahl (2003). The applicability of these methods for measuring context similarity in LBS-based CF should be carefully investigated and compared.

Incorporating context information into the CF process

In Adomavicius and Tuzhilin (2008), three approaches are proposed to incorporate context information into CF: 1) contextual pre-filtering: filter out irrelevant ratings (i.e., trajectories in our case) before using non-contextual CF method (i.e., the classical method); 2) contextual post-filtering: use the classical CF method, and then filter the recommendation results with context information; 3) contextual modeling: use context information directly inside the recommendation process. In Panniello et al. (2009), contextual pre-filtering and post-filtering approaches to include context are compared in an e-commerce portal. They found that post-filtering approach always reaches better performance than the pre-filtering approach.

Currently, the three approaches haven’t been applied to provide CaCF in LBS. How these three approaches can be combined with other key issues (i.e, annotating user profiles with context, measuring context similarity) to provide CaCF in LBS should be carefully investigated.

3. Methodology

In this section, we explore some methods to address the key challenges of incorporating context-aware CF into LBS (mentioned in section 2.4, i.e., identifying relevant context parameters, measuring context similarity, and making recommendations). Also methods of measuring user similarity based on trajectories are proposed.

3.1 Identifying Relevant Context Parameters

Context-dependent user profiles are important for context-aware recommendation. For annotating user profiles with context, a main question has to be answered: which context parameters are relevant and thus needed to be modeled.

We adopt the interactional perspective on context (Dourish 2004). Something is context (parameter) only if user's decision-making (e.g., choosing which POIs to visit), interaction with the system, or the behavior of the system depends on it, otherwise it is just a feature of the world (Winograd 2001). For example, the humidity of the room is a context parameter only if the adaptation of the interaction between human and the current system depends on it (or the behavior of the system depends on it), but otherwise it is just a feature of the world (Huang and Gartner 2009).

Based on this understanding, a two-stage method to identify relevant context parameters is designed:

1) A preliminary set of context parameters can be identified from literature or brain storming. And then, data about visitors' visit (i.e., trajectories) are collected and annotated with the preliminary set of context parameters.

2) The final set of context parameters can be created by refining the preliminary set according to the collected data. The basic strategy of refining is to analyze how some key aspects (e.g., the number of visited POIs, the length of the visit, and the duration of visit) of users' visits differ with different values of each context parameter in the preliminary set. If context parameter *c1* has *n* values, and the differences of the key aspects of visits are significant among these *n* values, then the current context parameter is relevant and thus needed to be modeled, otherwise it is irrelevant. F-test or T-test can be employed to measure the significance. For example, if context parameter *"weather"* has two values (e.g., *"sunny"* and *"rainy"*), and the difference between the key aspects (e.g., the number of visited POIs) of visits in *"sunny"* and the key aspects of visits in *"rainy"* is significant, then *"weather"* is relevant and thus needed to be modeled for CaCF, otherwise it is irrelevant.

It is important to note that we don't need to consider location as a relevant context parameter when annotating user profiles (trajectories) with context information. The reason is that location information is reflected in trajectories. A trajectory is a serial of different locations (i.e., POIs). Users' current location and location history are stored in their trajectories. Every POI has a location. When recommending a POI for current user, his/her current location (i.e., the current POI, obtained from his/her trajectory) is used to select relevant POIs which are *"close"* to the user (see step 1 of Method2_1 in section 3.4).

3.2 Measuring User Similarity (Taste)

For each user, a serial of POIs visited by him/her can be identified from his/her trajectory. Therefore, a simple user similarity measure is adopted. We measure similarity between two users by comparing POIs they visited.

Similar to Zheng et al. (2009), the visited popularity of a POI is considered when measuring the similarity between users. Two users accessed a POI visited by a few people might be more correlated than others who share a POI history accessed by many people. For instance, lots of people have visited the Great Wall, a well-known landmark in Beijing. However, it might not mean all these people are similar to one another. However, if two users visited a restaurant, which is not that famous, they might indeed share some similar preferences (Zheng et al. 2009).

Following is the proposed user similarity measure (taste).

$$SIM_{user}(a,b) = \frac{\sum_{p \in POIS_{a,b}} \frac{1}{F_p}}{\sqrt{(\sum_{p \in POIS_a} \frac{1}{F_p}) * (\sum_{p \in POIS_b} \frac{1}{F_p})}}$$

Where $POIS_a$ and $POIS_b$ are the set of visited POIs of user a and user b. $POIS_{a,b}$ is the set of POIs which are visited both by user a and user b. F_p is the visited popularity of POI p considering all the trajectories. $SIM_{user} \in [0,1]$, and 1 means these two users have identical preferences (taste).

3.3 Measuring Context Similarity

The similarity of the context (or situation) in which the trajectory is made with the current context of the active user (who asks for recommendations) determines the usefulness of this trajectory in recommending POIs for the active user. Two approaches are explored for measuring context similarity: Local-Global Approach (LGA), and Statistic-Based Approach (SBA).

3.3.1 Local_Global Approach (LGA)

The Local_Global Approach uses the local-global principle which is often employed to measure the similarity between complex case representations (i.e., context or situation in our case) consisting of attributes (i.e., context parameters in our case) with various different value types. According to this principle it is possible to decompose the entire similarity computation in a local part only considering local similarities between single attribute values, and a global part computing the global similarity for whole cases based on the local similarity assessments (Stahl 2003).

As a result, for measuring similarity between two contexts (situations), the following steps are applied:

1) For each relevant context parameter, calculate its local similarity. Context hierarchies may be employed for measuring local similarity. A simple local similarity with exact matching is applied in our experiment (in section 4). For example, the local similarity between *"weather: sunny"* and *"weather: rainy"* is 0, while the similarity between *"weather: sunny"* and *"weather: sunny"* is 1.

2) The global similarity between the two contexts can be calculated as the sum of the product of every local similarity and its importance weight. The important weights will be estimated and learned from the collected data. For example, we can evaluate several thousand parameterizations (e.g., varying the important weights), and use the best-performing one as the optimized weights.

3.3.2 Statistic-Based Approach (SBA)

The second approach adopts a machine-learning technique.

With the method proposed in section 3.1, relevant context parameters can be identified. By varying values for each parameter, all different kinds of situations can be identified. In the following, we propose an approach to measure the similarity between any two situations.

We assume that if visits in a situation (e.g., A) are similar to visits in another situation (e.g., B), then these two situations can be considered as similar. As a result, similarity between different contexts (situations) can be measured as some statistical metrics.

1) Measuring the distance of visits in situation A and visits in situation B:

$$Dist(A,B)=\sqrt{\frac{\sum_{p\in\bar{P}}\frac{1}{F_p}*(A_p-B_p)^2}{\sum_{p\in\bar{P}}\frac{1}{F_p}}}$$

$\bar{P}$ is the set of all POIs. A_p and B_p are the visit frequencies of POI p in situation A and B. F_p is the visited popularity of POI p considering all the trajectories.

2) Translating the distance measure into similarity measure:

Shepard (1987) proposes a universal law that distance and perceived similarity are related via an exponential function. As a result, the following context similarity measure is designed:

$$SIM_{conx}(A,B)=e^{-Dist(A,B)}$$

$SIM_{conx}\in[0,1]$, and 1 means these two situations are (nearly) identical.

With these two steps, similarity between any two situations can be calculated.

3.4 Making Recommendations

As mentioned in section 2.4, context information can be incorporated into CF by contextual pre-filtering, contextual post-filtering, and contextual modeling.

In this report, we mainly focus on contextual pre-filtering and contextual modelling. As a result, four kinds of CaCF methods can be designed (the current user is finishing the current POI *p*, and asking "which POI to visit next"):

1. ***Method2_1: Using Local-Global Approach and contextual pre-filtering (LGA_CP_CaCF)***

    1) Identifying users (i.e., trajectories) whose next POI after visiting *p* (the current POI) hasn't been visited by the current user.

    2) Filtering users (i.e., trajectories) whose context similarities with the current user don't exceed a threshold δ. Context similarity is measured by employing the LGA method proposed in section 3.3.1.

    3) For the results of step 2, identify the N most similar users. The proposed method in section 3.2 is used to calculate user similarities among different users.

    4) For the N most similar users, aggregating every similar user's next POI after visiting *p* (considering user similarity value).

    5) Selecting the POI with the highest predicted value, and recommending it to the current user.

2. ***Method2_2: Using Local-Global Approach and contextual modeling (LGA_CM_CaCF)***

    1) The same as step 1 in Method2_1 (LGA_CP_CaCF).

    2) For the results of step 1, identify the N most useful users. The usefulness is measured by considering both context similarity and user similarity.

$$Utility(a,b) = \lambda * SIM_{user}(a,b) + (1-\lambda) SIM_{conx}(C_a, C_b)$$

Where C_a and C_b are the context of user a and b. $SIM_{conx}(C_a, C_b)$ is calculated using the LGA method in section 3.3.1. $Utility \in [0,1]$, and 1 means these two trajectories share identical preferences and contexts.

3) For the N most useful users, aggregating every useful user's next POI after visiting *p* (considering usefulness value).

4) The same as step 5 in Method2_1 (LGA_CP_CaCF).

3. Method3_1: Using Statistic-Based Approach and contextual pre-filtering (SBA_CP_CaCF)

The steps are the same as steps in LGA_CP_CaCF, except that the context similarity in step 2 is measured by the SBA method.

4. Method3_2: Using Statistic-Based Approach and contextual modeling (SBA_CM_CaCF)

The steps are the same as steps in LGA_CM_CaCF, except that the context similarity in step 2 is measured by the SBA method.

With the above CaCF methods, context-aware recommendations can be provided in LBS.

4. Evaluation

In this section, we discuss our experimental evaluation. The data collection and processing are discussed in section 4.1. Section 4.2 employs the proposed method in section 3.1 to identify relevant context parameters. We describe the experiment setting in section 4.3. The evaluation and results are presented in section 4.4, and summarized in section 4.5.

4.1 Data Collection and Analysis

Thanks to the cooperation with Vienna Zoo (Schönbrunner Tiergarten), we collected trajectories in the zoo. GPS loggers (e.g., GPS travel recorder) were used to collect visitors' moving track. We encouraged visitors to carry those GPS loggers with them while walking through the zoo. Before they start, we recorded some additional information (e.g., context information) about the visitors, i.e., weather condition (sunny or rainy), age (>=45 or <45), time pressure (has two hours to stay in the zoo? Yes or No), annual ticket (Yes or No), first time in the Zoo (Yes or No), and companion with small children (Yes or No). In total, we collected 41 trajectories of all kinds of visitors in different weather conditions and different time of the day.

After the data collection, we converted the collected trajectories from the GPS loggers into *.gpx format. For every trajectory, we analyzed the data to identify the following information: Visited POIs and their orders, duration of the visit, and length of the visit. There is a small mountain in the zoo. As a result, whether the user visited the mountain or not is also considered.

In order to simplify the process of identifying the visited POIs from every trajectory, we defined 36 POIs (see Figure 1) in the zoo by considering the POI layout of the zoo and GPS accuracies. When a user's stop (where he/she has stayed in a certain distance threshold over a time period, in our case, distance threshold is set as 10m, and time threshold is set as 45s) is near/within a defined POI, the user is considered to have been visited the POI. As a result, a serial of POIs which users visited can be identified.

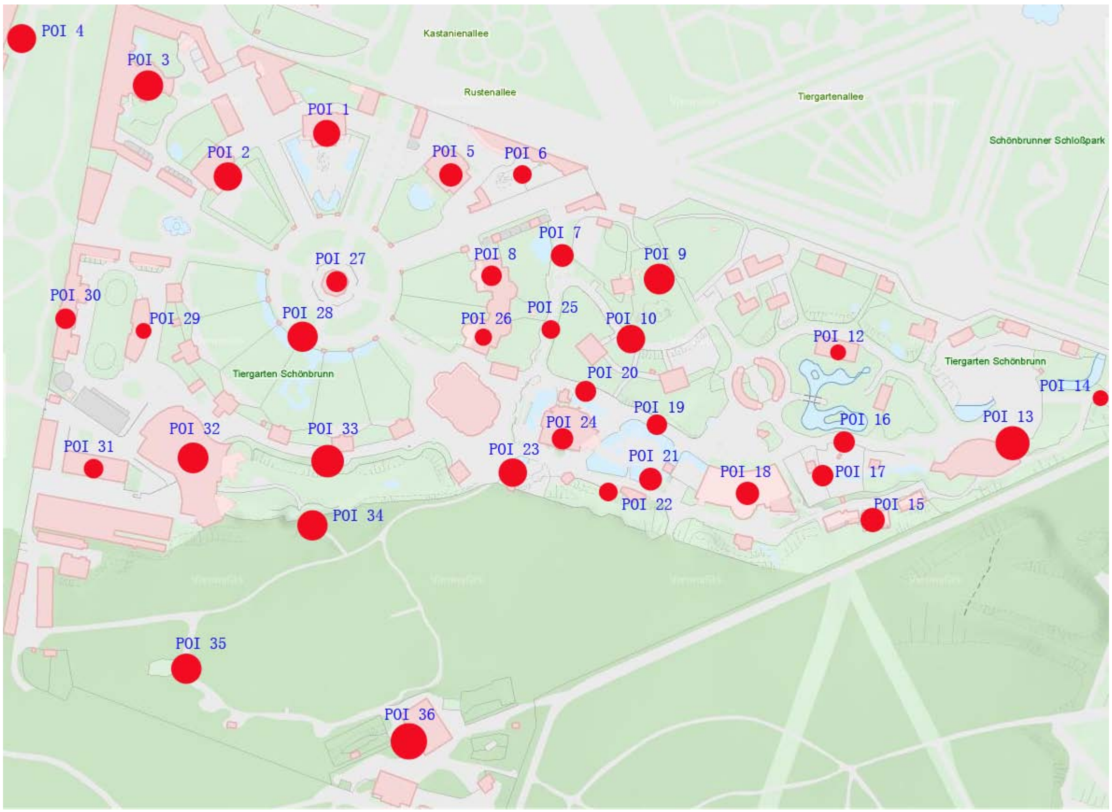

Figure 1 the defined POIs in Vienna Zoo

As a result, together with the recorded context information, for every collected trajectory, the following information is modeling:

<ID, visited POIs and their orders, number of visited POIs, length of visit, duration of visit, whether the user visited the mountain or not, age, first time in the Zoo, companion with small children, time pressure, annual ticket, weather >

Some statistical information about the collected data is shows in Table 1.

Table 1 some statistical information about the collected data

	Number of visited POIs	Length of visit	Duration of visit
Mean	13.90244	3.487804878	2.24
Standard Deviation	4.02371	2.654919502	1.444972

4.2 Identifying Relevant Context Parameters

The recorded context information (i.e., < “age”, “first time in the Zoo”, “companion with small children”, “time pressure”, “annual ticket”, “weather”>) can be viewed as the preliminary set of context parameters. In the following, we apply the proposed method in section 3.1 to identify relevant context parameters from this preliminary set.

We mainly compare the follow key aspects of visits among different situations: number of visited POIs, length of visit, duration of visit, and frequency of visiting the mountain.

Figure 2-5 depict how these aspects differ among different conditions for each context parameter.

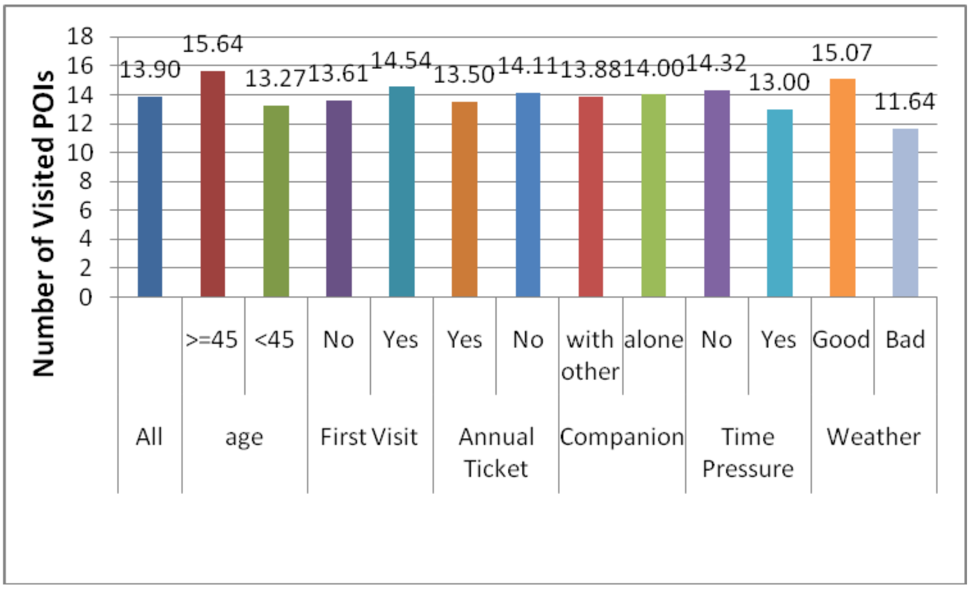

Figure 2 how numbers of visited POIs differ among different conditions for each context parameter

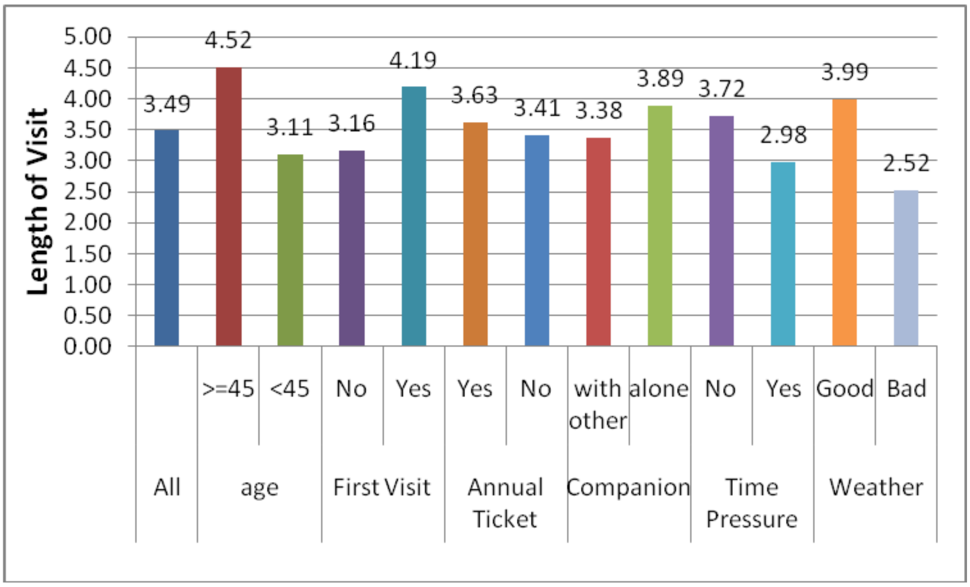

Figure 3 how length of visits differ among different conditions for each context parameter

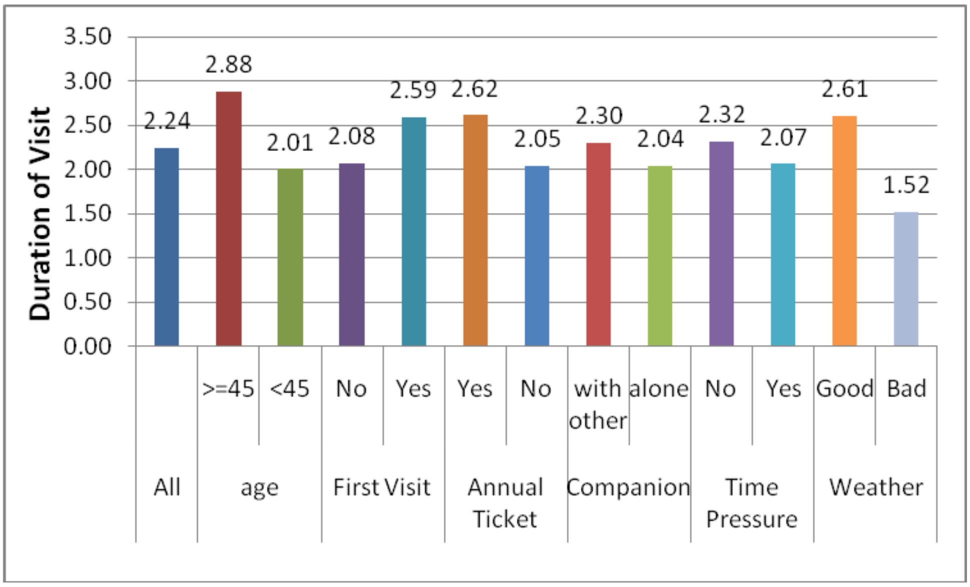

Figure 4 how durations of visits differ among different conditions for each context parameter

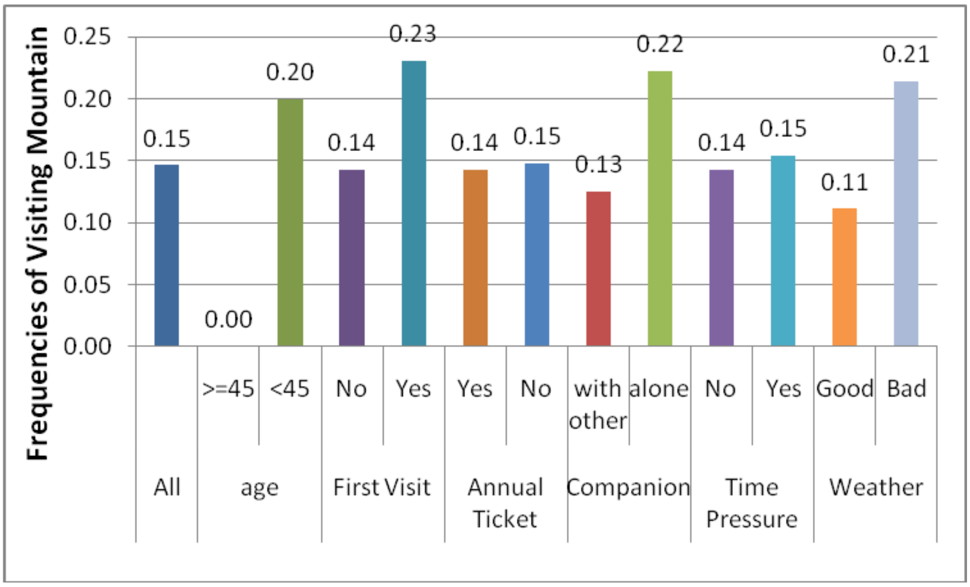

Figure 5 how frequencies of visiting the mountain differ among different conditions for each context parameter

From Figure 2-5, the differences among different conditions for “annual ticket” and “time pressure” are less obvious. In order to provide a combined view, we define a combined value:

$$Comb = \frac{diff_{num\_POIs}}{Aver_{num\_POIs}} + \frac{diff_{length}}{Aver_{length}} + \frac{diff_{duration}}{Aver_{duration}} + \frac{diff_{mountain}}{Aver_{mountain}}$$

Figure 6 shows the combined view about how visits differ among different conditions for each context parameter.

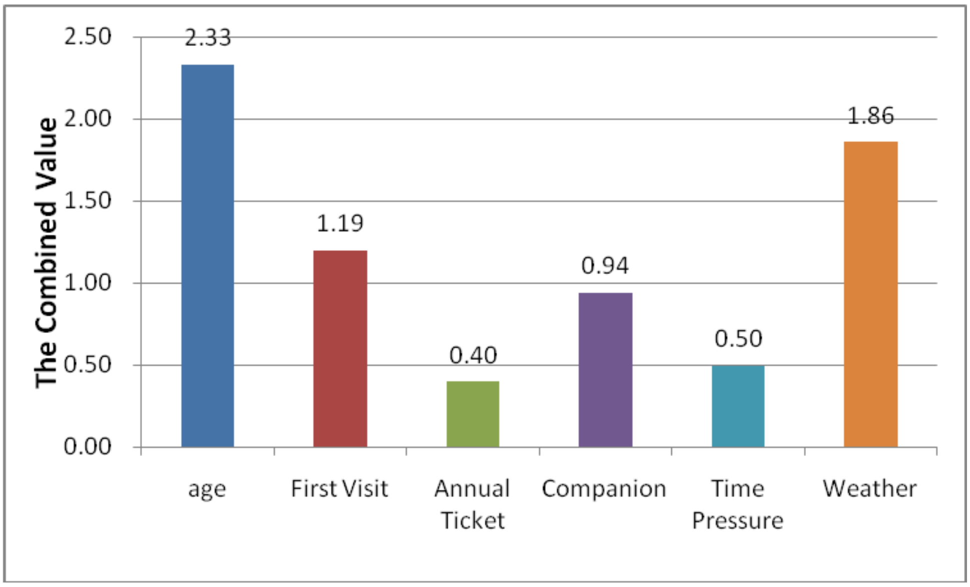

Figure 6 how visits differ among different conditions for each context parameter

From Figure 6, the context parameters ordered by relevances are: “age”, “weather”, “first visit”, “companion”, “time pressure”, and “annual ticket”.

In order to test whether the differences among different conditions for each context parameter are significant, we employ the F-test. Also due to the small size of our dataset, we consider $p \geq 0.15$ to indicate a lack of statistical significance.

Table 2 results of F-Test

	Number of visited POIs	Length of visit	Duration of visit
Age (>=45, <45)	2.926083***(p<0.15)***	2.340225***(p<0.15)***	3.06904***(p<0.15)***
First visit (Yes, No)	0.469311(p>0.15)	1.352194(p>0.15)	1.148168(p>0.15)
Year Ticket (Yes, No)	0.208456(p>0.15)	0.058358(p>0.15)	1.446467(p>0.15)
Companion (Yes, No)	0.006611(p>0.15)	0.258295(p>0.15)	0.224527(p>0.15)
Time Pressure(Yes, No)	0.956486(p>0.15)	0.678301(p>0.15)	0.27498(p>0.15)
Weather (Sunny, rainy)	7.852842***(p<0.15)***	11.02599***(p<0.15)***	19.84238***(p<0.15)***

The F-Test indicates that the variances (i.e., number of visited POIs, length of visit, and duration of visit) of the different groups of “age” are significantly different. Also the variances of the different groups of “weather” are significantly different. As a result, only “age” and “weather” are considered as relevant context parameters, and taken as the final set of context parameters.

In section 4.4, we provide an evaluation to test this decision.

4.3 Experiment Setting

We use the dataset described in section 4.1 to evaluate the predictive performance of the proposed CaCF methods: LGA_CP_CaCF, LGA_CM_CaCF, SBA_CP_CaCF, and SBA_CM_CaCF. In order to experimentally study whether including context information in a CF can improve the recommendation performance, we also implement a non-contextual CF method (non_CaCF, i.e., LGA_CP_CaCF ignoring step 2).

Due to the small size of our dataset, we use leave-one-out validation, i.e., we train our prediction models on 40 of the 41 visitors in our dataset, and test them on the remaining visitor (the active visitor). We use accuracy to evaluate the performance of the CaCF methods, and accuracy is defined as the ratio of the number of corrected recommedations (i.e., the predicited POI is actually viewed immediately by the active visitor) and the number of recommendation processes (i.e., 41 in our dataset).

In order to identify optimized values for different parameters in the proposed CaCF methods, we evaluate several thousand parameterisations (e.g., varying the important weights and the threshold in LGA_CP_CaCF), and use the best-performing one for our final experiments.

Two kinds of evaluations are performed. The first evaluation studies how the predictive performances of the proposed CaCF methods (LGA_CP_CaCF, LGA_CM_CaCF, SBA_CP_CaCF, and SBA_CM_CaCF) differ among different sets of context parameters. In additional to the proposed set of context parameters *“<age, weather>”*, 6 sets of context parameters are identified by considering the importances of parameters (see section 4.2): *“<age>”, “<weather>”, “<age, first visit, weather>”, “<age, first visit, companion, weather>”, “<age, first visit, companion, time pressure, weather>”*, and *“<age, first visit, annual ticket, companion, time pressure, weather>”*. The proposed CaCF methods using different sets of context parametres will be compared when predicting for the last POI of every visit (trajectory). Aslo non_CaCF is implemented as a benchmark. This evaluation is very useful for testing the effectiveness of the method proposed in section 3.1(identifying relevant context parameters).

The second evaluation focuses on how the predictive performances of the proposed CaCF methods differ when prediciting POIs at different places of a visit (i.e., the 1st last, the 2nd last, the 3rd last, the 4th last, and the 5th last). *"<age, weather>"* is used as the set of context parameters. non_CaCF is also implemented as a benchmark. This 2nd evaluation can help us to answer the following questions:

1) Does including context information in a CF for LBS improve the CF predictive performance (context-aware CF vs. non-contextual CF)?

2) How do the predictive performances of the proposed methods change when predicting POIs at different places of the visit?

4.4 Results

Figure 7 shows the results of how the predictive performances of the proposed CaCF methods change among different sets of context parameters (when predicting for the last POI of every trajectory).

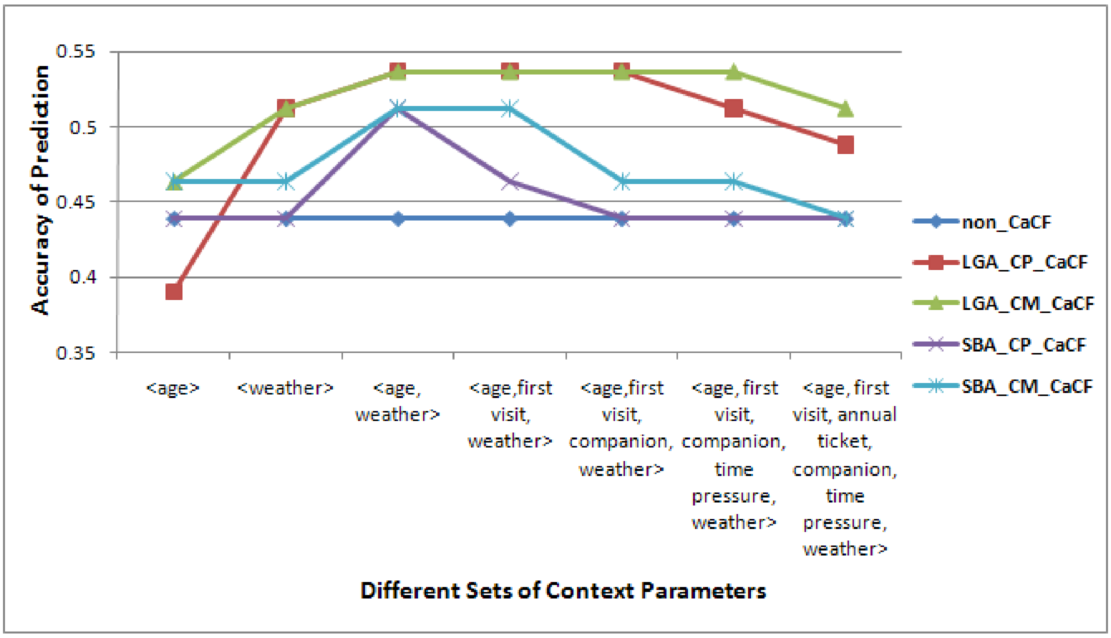

Figure 7 the predictive performances of the proposed CaCF methods change among different sets of context parameters (when predicting for the last POI of every trajectory)

Using different sets of context parameters: Figure 7 shows that the accuracy of CaCF methods (i.e., LGA_CP_CaCF, LGA_CM_CaCF, SBA_CP_CaCF, and SBA_CM_CaCF) using *“<age, weather>”* (i.e., the proposed set of context parameters in section 4.2) is *at least as high as* the accuracy of CaCF methods using other sets of context parameters. Specifically, the difference between the performance of all CaCF methods using *“<age, weather>”* and that of all CaCF methods using only 1 context parameter (i.e., *“<age>”*, and *“<weather>”)* is statistically significant ($p=0.003<0.15$ for *“<age>”*, $p=0.07<0.15$ for *“<weather>”*). This means that compared to the CaCF methods using only 1 context parameter, the CaCF methods using the proposed set of context parameters (“<age, weather>”) perform considerably better.

In the meanwhile, It is also important to note that incorporating more context parameters into the CF process doesn’t mean improvement of performance. This can be explained by the increasing difficulty of developing an accurate context similarity measure when using more context parameters, and the increasing demand of more data.

As a result, the following conclusion can be drawn:

1) Choose the suitable set of context parameters is very important for CaCF.
2) The proposed method to identify relevant context parameters is feasible, and using the proposed *“<age, weather>”* can achive a higher accuracy for all the designed CaCF methods.

Contextual Modelling vs. Contextual Pre-filtering when using different sets of context parameters: From Figure 7, the performances of different CaCF methods using different sets of context parameters can be compared. For all different sets of context parameters, the overall performance of LGA_CM_CaCF is at least as good as the performance of LGA_CP_CaCF. However, the difference between performance of LGA_CM_CaCF and that of LGA_CP_CaCF is not statistically significant ($p=0.45>0.15$). Similar results can be obtained when comparing SBA_CM_CaCF and SBA_CP_CaCF ($p=0.18>0.15$).

In conclusion, for different sets of context parameters, the performances of contextual modelling approaches (e.g., LGA_CM_CaCF, and SBA_CM_CaCF) are at least as good as the performances of contextual pre-filtering approaches(e.g., LGA_CP_CaCF, and SBA_CP_CaCF).

Local-Global Approach vs. Statistic-Based Approach when using different sets of context parameters: For all different sets of context parameters(except “<age>”), the performances of LGA (i.e., LGA_CP_CaCF and LGA_CM_CaCF) are considerably better than the performances of SBA (i.e., SBA_CP_CaCF and SBA_CM_CaCF). And differences between predictive performances are statistically significant ($p=0.05<0.15$ for LGA_CP_CaCF and SBA_CP_CaCF, $p=0.01<0.15$ for LGA_CP_CaCF and SBA_CM_CaCF, $p=0.0002<0.15$ for LGA_CM_CaCF and SBA_CP_CaCF, and $p=0.0025<0.15$ for LGA_CM_CaCF and SBA_CM_CaCF).

Context-aware CF vs. non-contextual CF when using different sets of context parameters: For all different sets of context parameters(except LGA_CP_CaCF using“<age>”), the performances of CaCF methods (i.e., LGA_CP_CaCF, LGA_CM_CaCF, SBA_CP_CaCF, SBA_CM_CaCF) are considerably better than the performances of non-contextual CF method (i.e., non_CaCF). The difference of performance of non_CaCF vs. that of LGA_CP_CaCF, performance of non_CaCF vs. that of LGA_CM_CaCF, and performance of non_CaCF and that of SBA_CM_CaCF, are statiscally significant ($p=1.59E\text{-}6<0.15$, $p=8.37E\text{-}9<0.15$, and $p=0.013<0.15$). However, the difference between performance of non_CaCF and that of SBA_CP_CaCF is not statistically significant ($p=0.21>0.15$).

Figure 8 shows the results of how the predictive performances of the proposed CaCF methods change when predicting POIs at different places of a visit (i.e., the 1^{st} last, the 2^{nd} last, the 3^{rd} last, the 4^{th} last, and the 5^{th} last). The proposed set of context parameters *(“<age, weather>”*) is employed.

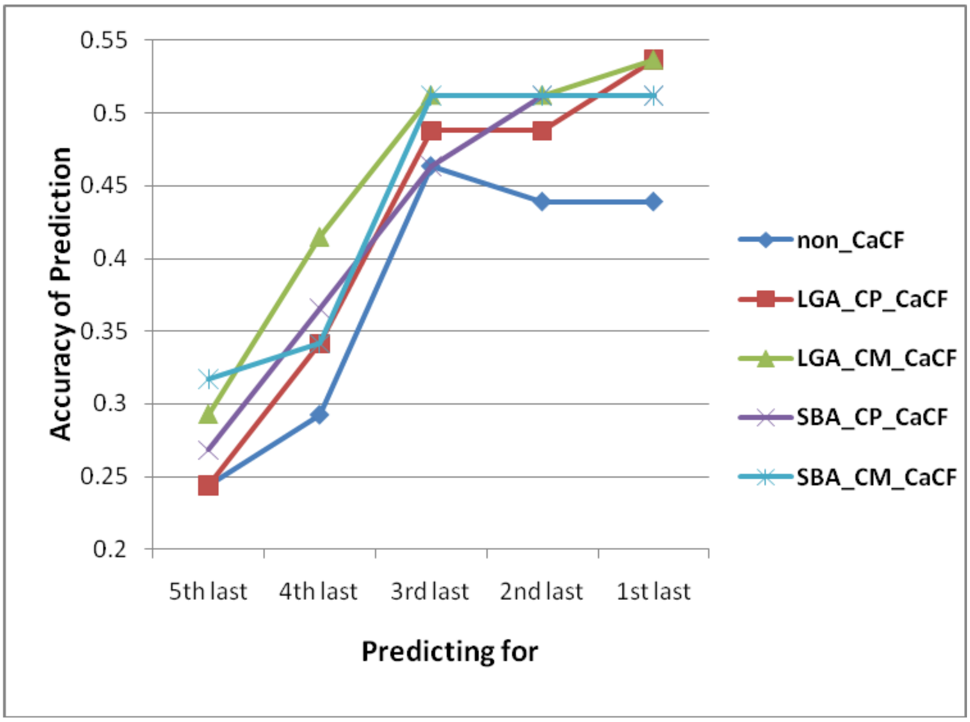

Figure 8 the predictive performances of the proposed CaCF methods change when prediciting POIs at different places of a visit (using *“<age, weather>”* as context parameters)

Predicting POIs at different places of a visit: Different places (positions) reflect the different amount of information available about a visitor. Figure 8 shows a upwards trend for the accuracy of all CaCF methods and non-contextual methods when the positions of the predicted POI increase. The difference of predictive performance for “5th last” vs. that for “4th last”, and predictive performance for “4th last” vs. that for “3rd last”, are statistically significant (*p*=0.01<0.15, and *p*=0.0003<0.15). The difference of predictive performance for “3rd last” vs. that for “2nd last”, and predictive performance for “2nd last” vs. that for “1st last”, are not statistically significant (*p*=0.79>0.15, and *p*=0.54>0.15).

In conclusion, with more information about a visitor available, the predictive performances of all the proposed CaCF methods are improving.

Contextual modelling vs. Contextual pre-filtering when prediciting POIs at different places of a visit: When prediciting POIs at different places, the overall performance of LGA_CM_CaCF is at least as good as the performance of LGA_CP_CaCF. However, the difference is not statistically significant (*p*=0.64>0.15). Similar results can be obtained when comparing SBA_CM_CaCF and SBA_CP_CaCF (*p*=0.83>0.15).

In conclusion, when prediciting POIs at different places, the overall performance of contextual modelling approach (e.g., LGA_CM_CaCF, and SBA_CM_CaCF) is at least as good as the performance of contextual pre-filtering approach(e.g., LGA_CP_CaCF, and SBA_CP_CaCF). An explanation for this would be: The latter suffers from the problems of sparsity as lots of trajectories are filtering out, while in contextual modelling, more users (trajectories) are involved in making recommendations.

Local-Global Approach vs. Statistic-Based Approach when prediciting POIs at different places of a visit: a clear conclusion comparing the performacnes of LGA and SBA can't be obtained from Figure 8.

Context-aware CF vs. non-contextual CF when prediciting POIs at different places of a visit: when prediciting POIs at different places, the performances of CaCF methods (i.e., LGA_CP_CaCF, LGA_CM_CaCF, SBA_CP_CaCF, SBA_CM_CaCF) are considerably better than the performances of non-contextual CF method (i.e., non_CaCF). However, all the differences are not statistically significant (p=0.65>0.15 for non_CaCF and LGA_CP_CaCF, p=0.33>0.15 for non_CaCF and LGA_CM_CaCF, p=0.58>0.15 for non_CaCF and SBA_CP_CaCF, and p=0.44>0.15 for non_CaCF and SBA_CM_CaCF). It is proposed that when more trajectories are available, the predictive performance of CaCF methods will be improved and will have a significant differences with non-contextual CF method.

In conclusion, the proposed CaCF methods provide better predictive performance than non-contextual CF. Among different CaCF methods, LGA_CM_CaCF provides the best predictive performance. However, all the differences in performances are not statistically significant.

4.5 Summary

In summary, the main findings of the experiment are as follows.

1) When including context information in the CF process, choosing a suitable set of relevant context parameters is very important and may affect the predictive performance.

2) The proposed method to identify relevant context parameters is feasible and useful, and using the proposed *"<age, weather>"* can achive a higher accuracy for all the designed CaCF methods.

3) The proposed CaCF methods provide better performance than non-contextual CF, that means including context information in a CF for LBS can improve the predictive performance.

4) Among different CaCF methods, LGA_CM_CaCF provides the best predictive performance.

5) The predictive performance of contextual modelling approach is at least as good as the performance of contextual pre-filtering approach.

5. Conclusions and Future Work

In this report, methods of introducing context-aware collaborative filtering (CaCF) into Location Based Services (LBS) are proposed. To be more specific, CaCF methods are applied on the high available spatio-temporal trajectories to enhance visitors with context-aware POI recommendations in LBS applications.

The main contributions are as follows:

1) Key issues of CaCF in LBS applications are identified

2) A two-stage method is designed to identify relevant context parameters for CaCF.

3) Two different approaches (i.e., Local-Global Approach and Statistic-Based Approach) are proposed to measure context similarity.

4) In considering two different ways (i.e., contextual pre-filtering, and contextual modeling) of incorporating context information into the CF process, four CaCF methods are designed for LBS applications: LGA_CP_CaCF, LGA_CM_CaCF, SBA_CP_CaCF, and SBA_CM_CaCF.

5) Experimental studies are designed to evaluate the proposed methods. The results of

the experiment show that the proposed CaCF methods are feasible and useful for providing context-aware recommendation in LBS applications.

From the experiment, following conclusion can be drawn: including context information in the CF process can improve the predictive performances.

Our next step is to collect more trajectories data in both outdoor (i.e., Vienna Zoo) and indoor to evaluate the proposed methods. We propose that with more trajectories available, the predictive performance of CaCF methods will be improved and will have a significant differences with non-contextual CF method. Also a clear conclusion comparing the performacnes of LGA and SBA can be made. We are also interested in exploring a more complex user simiarlty measure in considering spatio-temporal behavior to provide more accurate results.

Acknowledgement

Haosheng Huang is a recipient of the Marshall Plan Scholarship of the Austrian Marshall Plan Foundation. With this scholarship, he visited the University of California, Santa Barbara from February 2010 to July 2010. Special thanks to Prof. Georg Gartner at the Vienna University of Technology, and Prof. Martin Raubal at the University of California, Santa Barbara for their supervision. Huang also gratefully acknowledges the contributions of students (Martin Fleischhacker, Manuela Handl, Connie Kwok, Klemens Lagler, and Bernd Moser) from the LBS class (VU 126.099, 2010SS) for collecting the data in Vienna Zoo.

Reference

Adomavicius, G., Sankaranarayanan, R., Sen, S. and Tuzhilin, A. (2005): Incorporating contextual information in recommender systems using a multidimensional approach. ACM Transactions on Information Systems, 23 (1), 103-145.

Adomavicius, G. and Tuzhilin, A. (2008): Context-aware recommender systems. In: RecSys'08, http://ids.csom.umn.edu/faculty/gedas/talks/RecSys2008-tutorial.pdf, Accessed on 01.2010.

Baus, J., Kray, C. and Krüger, A. (2001): Visualization of route descriptions in a resource-adaptive navigation aid. Cognitive Processing, 2 (2-3):323-345.

Baus, J., Chevherst, K. and Kray, C. (2005): A survey of map-based mobile guides. In: Meng, L., Zipf, A.

and Reichenbacher, T. (eds.): Map-Based Mobile Services, Springer, Berlin, 197-216.

Bohnert, F., Zukerman, I., Berkovsky, S., Baldwin, T. and Sonenberg, L. (2008): Using collaborative models to adaptively predict visitor locations in museums. In: Proc. of the Fifth Intl. Conf. on Adaptive Hypermedia and Adaptive Web-Based Systems (AH-08).

Chen, A. (2005): Context-aware collaborative filtering system: Predicting the user's preference in the ubiquitous computing environment. In: Strang, T. and Linnhoff-Popien, C. (eds.): Proc. of LoCA 2005, Springer LNCS 3479, 244-253.

Cheverst, K., Davies, N., Mitchell, K., Friday, A. and Efstratiou, C.(2000): Developing a context-aware electronic tour guide: Some issues and experiences. In: Proc. of the Conference on Human Factors in Computing Systems (CHI 2000), 17-24.

de Spindler, A., Norrie, M., Grossniklaus, M. and Signer, B. (2006): Spatio-temporal proximity as a basis for collaborative filtering in mobile environments. In: Proc. of UMICS'06, Luxemburg, June 5-9, 2006.

Dourish, P. (2004): What we talk about when we talk about context. Personal and Ubiquitous Computing, 8(1), 19–30.

Gonotti, F., Nanni, M., Pedreschi, D. and Pinelli, F. (2007): Trajectory pattern mining. In Proc. of KDD’07, ACM Press, pp. 330-339.

Horozov, T., Narasimhan, N. and Vasudevan, V. (2006): Using location for personalized POI recommendations in mobile environments. In: Proc. of the International Symposium on Applications on Internet 2006, 124-129.

Huang, H. and Gartner, G. (2009): Using Activity Theory to Identify Relevant Context Parameters. In Gartner, G. and Rehrl, K. (eds.): Location Based Services and TeleCartography II - From Sensor Fusion to Context Models, Springer Verlag Berlin Heidelberg.

Li, Q., Zheng, Y., Xie, X., Chen, Y., Liu, W, and Ma, W. (2008): Mining user similarity based on location history. In Proc. of GIS’08 (Santa Ana, CA, Nov.2008). ACM Press, pp.298-307

Li, L., Lee, F., Chen, Y., and Cheng, C. (2009): A multi-stage collaborative filtering approach for mobile recommendation. In Proc. of the 3rd International Conference on Ubiquitous Information Management and Communication, January 15-16, 2009, Suwon, Korea.

Kray, C. (2003): Situated interaction on spatial topics. PhD thesis, Saarland University, Germany.

Krüger, A., Baus, J., Heckmann, D., Kruppa, M. and Wasinger, R. (2007): Adaptive mobile guides. In: Brusilovsky, P., Kobsa, A. and Nejdl, W. (eds): The adaptive Web, Springer, Berlin, 521-549.

Panniello, U., Gorgoglione, M. and Palmisano, C. (2009): Comparing pre-filtering and post-filtering approach in a collaborative contextual recommender system: An application to E-Commerce. In: Noia, T. and Buccafurri, F. (eds.): Proc. of EC-Web 2009, Springer LNCS 5692, 348-359.

Raper, J., Gartner, G., Karimi, H. and Rizos, C. (2007): Applications of location-based services: a selected review. Journal of Location Based Services, 1(2), 89-111.

Resnick, P. and Varian, H. (1997): Recommender systems. Communications of the ACM, 40(3), 50-58.

Schmidt, A., Beigl, M. and Gellersen, H. (1999): There is more to context than location. Computers and Graphics, 23(6), 893–901.

Schwinger, W., Grün, C., Pröll, B., Retschitzegger, W. and Schauerhuber, A. (2005): Context-awareness in mobile tourism guides – A comprehensive survey. Technical Report, Johannes Kepler University

Linz.

Shepard, R. (1987): Toward a universal law of generalization for psychological science. Science, 237:1317-1323.

Stahl, A. (2003): Learning of knowledge-intensive similarity measures in case-based reasoning. PhD thesis, University of Kaiserslautern.

Takeuchi, Y. and Sugimoto, M. (2006): CityVoyager: An outdoor recommendation system based on user location history. In: J. Ma et al. (eds.): UIC 2006. Springer Berlin/Heidelberg, 625–636.

van Setten, M., Pokraev, S. and Koolwaaij J. (2004): Context-aware recommendations in the mobile visitor application COMPASS. In: Nejdl, W. and De Bra, P. (eds.): Proc. of AH 2004, Springer LNCS 3137.

Wexelblat, A. (1999): Footprints: Interaction history for digital objects. PhD thesis, MIT Program in Media Arts & Sciences.

Wiesenhofer, S., Feiertag, H., Ray, M., Paletta, L., Luley, P., Almer, A., Schardt, M., Ringert, J., Beyer, P. (2007): Mobile city explorer: An innovative GPS and camera phone based travel assistant for city visitors. In: Location Based Services and TeleCartography, Springer, 557-573.

Winograd, T. (2001): Architectures for context. Human-Computer Interaction, 16(2/4), 401-419.

Zheng, Y., Liu, L., Wang, L. and Xie, X. (2008): Learning transportation mode from raw GPS data for geographic applications on the Web. In: Proc. WWW 2008, ACM Press, pp. 247-256.

Zheng, Y., Chen, Y., Xie, X. and Ma, W. (2009): GeoLife2.0: A Location-Based Social Networking Service. In: Proc. of the 2009 Tenth International Conference on Mobile Data Management: Systems, Services and Middleware, p.357-358, May 18-20, 2009

Zheng, Y., Zhang, L., Xie, X. and Ma. W. (2009): Mining interesting locations and travel sequences from GPS trajectories. In: Proc. WWW 2009.