Knowledge-Based Systems 73 (2015) 97–110

Contents lists available at ScienceDirect

Knowledge-Based Systems

journal homepage: www.elsevier.com/locate/knosys

A Location-Item-Time sequential pattern mining algorithm for route recommendation

Chieh-Yuan Tsai [a,b,*], Bo-Han Lai [a]

[a] Department of Industrial Engineering and Management, Yuan-Ze University, Taiwan
[b] Innovation Center for Big Data and Digital Convergence, Yuan-Ze University, Taiwan

ARTICLE INFO

Article history:
Received 14 January 2014
Received in revised form 5 September 2014
Accepted 26 September 2014
Available online 5 October 2014

Keywords:
Recommendation systems
Sequential pattern
Sequence mining
Behavior computing
Theme park

ABSTRACT

To survive in a rapidly changing environment, theme parks need to provide high quality services in terms of visitor tastes and preferences. Understanding the spatial and temporal behavior of visitors could enhance the attraction management and geographical distribution for visitors. To fulfill the need, this research defined a Location-Item-Time (LIT) sequence to describe visitor's spatial and temporal behavior. Then, the Location-Item-Time PrefixSpan (LIT-PrefixSpan) mining algorithm is developed to discover frequent LIT sequential patterns. Next, the route suggestion procedure is proposed to retrieve suitable LIT sequential patterns for visitors under the constraints of their intended-visiting time, favorite regions, and favorite recreation facilities. A simplified theme park is used as an example to show the feasibility of the proposed system. The experimental results show that the system can help managers understand visitors' behavior and provide appropriate visiting experiences for visitors.

1. Introduction

A theme park is an aggregation of attractions including architecture, landscape, rides, shows, food services, costumed personnel and retail shops. Well-known examples include Disney World, Disneyland, Universal Studios and Six Flags. Although the theme park industry has enjoyed steady attendance growth in the past several decades, the theme park market has entered a mature stage and is no longer experiencing high growth [5,6]. To survive in a rapidly changing environment, theme parks need to provide high quality services in terms of visitor tastes and preferences. Understanding the spatial and temporal behavior of visitors could enhance the management of attractions and contribute to extending the geographical distribution of visitors within regions.

In the past decade, the recommendation technique has been regarded as a popular technique for providing a variety of products, services and items to customers in the tourism industry [4,7,13]. Personalized tourism services aim at helping users to find what they are looking for by comparing the user profile to reference characteristics. Wang et al. [19] presented semantic web technologies for providing personalized access to digital museum collections. Niaraki and Kim [12] proposed a generic ontology-based architecture using a multi-criteria decision making technique to design a personalized route planning system. Schiaffino and Amandi [14] developed an expert software agent in the tourism and travel domain, named Traveler. This agent combines collaborative filtering with content-based recommendations and demographic information about customers to make recommendations. García-Crespo et al. [3] presented the SPETA system, which uses knowledge of user's current location, preferences, as well as a history of past locations to provide the type of recommendation services that tourists expect from a real tour guide. Tsai and Lo [17] took previous popular visiting behaviors as the foundation and developed a sequential pattern based route suggestion system to generate personalized tours. Tsai and Chung [16] developed a route recommendation system that provides personalized visiting routes for tourist in theme parks that consider a set of visiting sequences. Based on the retrieved visiting behavior data and facility queuing situation, their system can generate a proper route suggestion for visitors.

The above recommendation systems have demonstrated themselves efficient tools by designing user interfaces that can smoothly interact with the environment, providing convenient information query tools, or suggesting a set of associated products (or services). However, three major problems are revealed. First, these systems simply return a set of suggested facilities (items) in a sequential order, but fail to illustrate the complete visiting path for visitors. For example, their systems might suggest a visitor visit items k_1, k_4, and k_8 in order (i.e., $k_1 \rightarrow k_4 \rightarrow k_8$). However, the actual path

* Corresponding author at: Department of Industrial Engineering and Management, Yuan-Ze University, Taiwan.
E-mail address: cytsai@saturn.yzu.edu.tw (C.-Y. Tsai).

http://dx.doi.org/10.1016/j.knosys.2014.09.012
0950-7051/

to complete the route should contain "by-pass items" such as $k_1 \rightarrow k_4 \rightarrow k_7 \rightarrow k_8$, $k_1 \rightarrow k_4 \rightarrow k_6 \rightarrow k_8$, or even $k_1 \rightarrow k_4 \rightarrow k_7 \rightarrow k_6 \rightarrow k_8$. Without providing complete path information, a visitor might get confused and spend much more time to finish the route. Second, previous systems seldom take the geographic constraints into consideration so that their suggested routes are often trivial and impractical. For example, previous studies might suggest visitor a long route $k_1 \rightarrow k_2 \rightarrow k_6 \rightarrow k_4 \rightarrow k_7 \rightarrow k_{10} \rightarrow k_8 \rightarrow k_{12}$. However, the route is trivial and hard-to-follow since k_1, k_2, and k_4 are in region A, k_6, k_7 and k_8 are in region B, and k_{10} and k_{12} in region C. In fact, a theme park consists of several regions where each region contains dozens of facilities and shops. It will be worthwhile to suggest a no-trivial suggestion such as A(k_1, k_4, k_2) → B(k_8, k_6) → C(k_{10}, k_{12}) for visitors. Third, previous studies seldom took the time constraints into consideration when they provided route suggestion for visitors. For example, previous systems simply suggest a route format such as $k_1 \rightarrow k_4 \rightarrow k_8$ for visitors. However, when time interval information between items are revealed, this route will be $k_1 \rightarrow (1\text{ h}) \rightarrow k_4 \rightarrow (1\text{ h}) \rightarrow k_8$. If the intended-visiting time for a visitor is 90 min, this suggestion is unacceptable since the visitor cannot finish the route on time. On the other hand, if intended-visiting time is 300 min, this suggestion is not suitable also. Without providing time interval between items in the suggestion, tourists are unsure whether she/he can complete the suggested route on time or not.

To solve the above problems, this research defines a Location-Item-Time (LIT) sequence to describe visitor's spatial and temporal behavior. To the best of our knowledge, this study is the first work to include location (region), item, and time-interval information simultaneously into a sequence. Then, the Location-Item-Time PrefixSpan (LIT-PrefixSpan) mining procedure is developed to discover frequent LIT sequential patterns. Finally, the route suggestion procedure is proposed to retrieve suitable LIT sequential patterns under the constraints of visitor's intended-visiting time, favorite regions and its related visiting time, favorite recreation facilities. This paper is organized as follows. Section 2 reviews previous works related to sequential pattern mining and suggestion. Section 3 introduces the framework of the proposed route recommendation system. Section 4 demonstrates a case to show the feasibility of the proposed system. Finally, Section 5 summarizes the conclusions and points out possible future directions.

2. Literature review

Yavas et al. [20] proposed a three-phase mobility prediction algorithm for the prediction of user movement in a mobile computing system. Their algorithm enables the system to allocate resource for users in an efficient manner, and to produce more accurate answers to location-dependent queries that refer to future positions of mobile users. Cho et al. [2] proposed a sequential rule-based recommendation method that considers the evolution of customers' purchase sequences. The purchase transaction records of a customer for a certain period are used to build a customer profile. Then, a collaboration-based system is in charge to find a set of customers, through calculating the correlations among customers profile. Tan et al. [15] proposed a new approach to build personalization recommendation system based on access sequential patterns, named Frequent Accessed Sequence Tree (FAS-Tree). All frequent access sequential patterns are compressed into FAS-Tree to save storage greatly. During personalization recommendation stage, it is only necessary to traverse sub paths of FAS-tree referring to page views in active window to find match patterns, without the need to generate association rules. Yun and Chen [21] developed a mining mobile sequential patterns algorithm to better reflect the customer usage patterns in the mobile commerce environment, which takes both the moving patterns (location) and purchase patterns (items) of customers into consideration. Tseng and Lin [18] proposed a novel data mining method, namely SMAP-Mine that can efficiently discover mobile users' sequential movement patterns associated with requested services. Through empirical evaluation under various simulation conditions, SMAP-Mine is shown to deliver excellent performance in terms of accuracy, execution efficiency and scalability. Meanwhile, the proposed prediction strategies are also verified to be effective in measurements of precision, hit ratio and applicability.

Li et al. [8] proposed a Multi-Stage Collaborative Filtering (MSCF) process to provide the location-aware event recommendation service in mobile environment. The first stage in MSCF performs the People-to-People Collaborative Filtering (P2P-CF), while the Event-to-Event Collaborative Filtering (E2E-CF) discovers the sequential rules of event-participation in the second stage. Liu and Chang [9] proposed a route recommendation system which guides the user through a series of locations. Their system used the methods of sequential pattern mining to extract popular route patterns from a large set of historical user's route records. Then, the system recommends routes by matching the user's current route with the set of popular route patterns. Liu et al. [10] proposed a novel hybrid recommendation approach that combines the segmentation-based sequential rule (SSR) method with the segmentation-based KNN-CF (SKCF) method. In order to enhance the quality of product recommendations, their method considers customers' purchase sequences over time and their purchase data for the current period. Hung and Peng [6] proposed a Regression-based approach for mining User Movement Patterns (RUMP). Large Sequence (LS) algorithm extracts the call detail records and Time Clustering (TC) algorithm determines the number of regression functions. Then, Movement Function (MF) algorithm generates the movement function representing user movement patterns of mobile users. Lu et al. [11] proposed a hybrid semantic recommendation approach which integrates item-based CF similarity with item-based semantic similarity techniques. The hybrid semantic recommendation approach has been implemented in an Intelligent Business Partner Locator recommendation system prototype named BizSeeker. Similarly, Zhang et al. [22] developed a hybrid recommendation approach which combines user-based and item-based collaborative filtering techniques with fuzzy set techniques and knowledge base for mobile product and service recommendation. It particularly implements the approach in a personalized recommender system for telecom products/services called FTCP-RS. Although the above sequential pattern algorithms are efficient in different environment, however, they did not take location, item, and time-interval information into consideration at the same time.

3. Research method

3.1. Environment assumption and system overview

Typically, a theme park is divided into several regions and each region contains a set of recreation facilities. It is assumed that each region is fully covered by RFID readers. In addition, RFID readers are installed in the entrance of each recreation facility, and entrance and exit of the park. When a visitor with a RFID tagged wristband enters a region or entrance of a facility, RFID readers record the RFID tag code, region id, facility id, and the time into a route database. The recording process continues until the visitor leaves the park. Let's take the layout in Fig. 1 as an example. At timestamp t_1, a visitor passes the entrance k_{11} of the park in region B. Then, she moves to region A at timestamp t_2, region F at timestamp t_3, and region G at timestamp t_4. In region G, she takes facility k_1. After that, she moves to region K at timestamp t_5, region O at timestamp t_6. In region O, she takes facilities k_2 and k_3. The recording process continues until she leaves the park from the exit