Computers & Industrial Engineering 106 (2017) 192–204

Contents lists available at ScienceDirect

Computers & Industrial Engineering

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

A shopping behavior prediction system: Considering moving patterns and product characteristics

Chieh-Yuan Tsai [a,b,*], Ming-Hong Li [a], R.J. Kuo [c]

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

ARTICLE INFO

Article history:
Received 28 April 2016
Received in revised form 3 February 2017
Accepted 3 February 2017
Available online 9 February 2017

Keywords:
Shopping behavior prediction
High-utility sequential patterns
Location based service
Mobile commerce

ABSTRACT

In recent years, the development of location determination technologies such as GPS, Wi-Fi, and RFID have made it possible to collect locational data for moving customers, and many behavior prediction and recommendation systems have been proposed based on moving path and purchase transactions. However, these systems do not take item-specific profit margins into consideration. In addition, few such proposals accounted for location similarity and item similarity. To address these issues, a shopping behavior prediction (SBP) system is proposed consisting of a behavior mining module, a similarity inference module, and a behavior prediction module. The behavior mining module can discover high-utility mobile sequential patterns (UMSPs) using the $UMSP_L$ algorithm. In the similarity inference module, store-to-store (StoS) similarities and item-to-item (ItoI) similarities are derived by a proposed similarity inference algorithm. When evaluating StoS and ItoI similarities, the quantities of items purchased are considered. Finally, based on UMSPs, StoS similarities, and ItoI similarities, the behavior prediction module generates a list of shopping suggestions for the target user.

1. Introduction

Customer relationship management (CRM) is a critical strategy for retail stores. To understand the purchase behavior of customers in stores, massive amounts of point of sale (POS) data have been accumulated for use in various CRM models to elicit insight into market segmentation, customer retention, shelf management, and one-to-one in-store promotion (Klabjan & Pei, 2011; Ngai, Xiu, & Chau, 2009; Tsai & Huang, 2015; Van den Poel, De Schamphelaere, & Wets, 2004). In recent years, the development and widespread deployment of location acquisition technology such as GPS, Wi-Fi, and RFID have made it possible to collect data on various aspects of customer mobility (Campos, Lovisolo, & de Campos, 2014; Larson, Bradlow, & Fader, 2005; Liao & Lin, 2007). Combined moving logs and POS payment records can help store managers identify which product zones in the store were visited, how long the person walked around the store, how long the person thought about a purchase, and when the purchase was made (Nakahara & Yada, 2012). For example, a customer shopping sequence <{<A; nail polish> <C; pants>}; ABC> means a customer moved through the path of locations A, B, and C, and bought nail polish and pants in locations A and C, respectively.

To provide better CRM service, some behavior prediction and recommendation systems have been proposed based on moving path and purchase transaction of customers (Lu, Lee, & Tseng, 2012; Lu & Tseng, 2009; Tseng, Lu, Huang, & Distributed Systems – ICPADS, 2007; Yavas, Katsaros, Ulusoy, & Manolopoulos, 2005) These prediction and recommendation systems have been shown to be useful in predicting customer behavior through consideration of moving path and purchase items. However, two major problems are found when these approaches are applied to actual retail store environments. First, previous behavior prediction/recommendation systems did not take the profit margin of specific items into consideration when deriving behavior patterns. This makes it difficult to determine more valuable behavior patterns (or high utility patterns). For example, assume two customer behaviors S_1 = <{<A; pants; 3> <H; nail polish; 6>}; AEFGH> and S_2 = <{<A; pants; 3> <H; diamond rings; 1>}; AEFGH>. Behavior S_2 appears much less frequently than behavior S_1, and thus previous systems are unable to discover S_2. However, store managers should be very interested in S_2 since the profit margin for diamond rings is much higher than for nail polish. If we can take the utility of each product item into consideration when deriving behavior patterns, the prediction/recommendation result should be more valuable for retailers.

* Corresponding author.
E-mail address: cytsai@saturn.yzu.edu.tw (C.-Y. Tsai).

http://dx.doi.org/10.1016/j.cie.2017.02.004
0360-8352/

Second, before generating suggestion results, similarity among behavior patterns should be carefully evaluated. Most previous studies considered a pattern as a string and evaluated similarity between patterns by Levenshtein distance (or edit distance). Few such studies accounted for location similarity or item similarity (Lu et al., 2012). Typically, consumers purchased similar items in similar stores (locations). Stores that sell similar items are therefore more likely to be considered similar. On the other hand, similar items are often sold in similar stores. Thus, items are considered to be similar if they are sold in similar stores. If store-to-store (StoS) similarity and item-to-item (ItoI) similarity can be derived and used when evaluating similarity between behavior patterns, it will produce more helpful suggestion results. In addition, the quantity of items sold plays an important role in evaluating StoS similarity and ItoI similarity. For example, hamburgers, French fries, and fried chicken can be bought in StoreA, StoreB, and StoreC. The daily sales volume for hamburgers, French fries, and fried chicken is [550, 50, 600] for StoreA, [100, 650, 120] for StoreB, and [650, 60, 550] for StoreC. Without considering quantity of a product sold for each store, previous study simply claims that the store similarity among StoreA, StoreB, and StoreC are the same since the three stores sold the same product items. However, the similarity between StoreA and StoreC should be higher than the similarity between StoreA and StoreB if sales volume for each product is considered. If we know that a person likes to visit StoreA but StoreA is not available now, StoreC should be a better alternative than StoreB.

To address these issues, this research proposes a shopping behavior prediction (SBP) system consisting of a mobile transaction sequence database, a behavior mining module, a similarity inference module, and a behavior prediction module. The remainder of this paper is organized as follows. Related works are reviewed in Section 2. Section 3 formally defines the research problem and provides a detailed description of the proposed approach. Section 4 presents an empirical evaluation and describes a set of experiments. Finally, conclusions are drawn and future work directions are proposed in Section 5.

2. Literature review

The moving/mobile behavior prediction can be roughly divided into two categories (Lu et al., 2012). The first category is time series-based prediction that can be further divided into two types which are linear models and nonlinear models. The nonlinear models consider an object's movements by more sophisticated regression functions, so that their prediction accuracies are higher than those of the linear models (Tao, Faloutsos, Papadias, & Liu, 2004). The second category is pattern-based prediction. In recent years, increased attention has focused on integration of pattern mining and rule matching techniques in moving/mobile environments to increase prediction accuracy (Tsai, Lo, & Lin, 2011; Vu, Ryu, & Park, 2009). In this category, the moving patterns are generated by specific pattern mining algorithms and represented as a set of sequential patterns or rules. Based on the patterns, variant prediction methods are developed to predict the next possible moving behavior of a user. Thus, the following discussion concentrates on the research related to pattern-based prediction category.

Jeung, Liu, Shen, and Zhou (2004) proposed the hybrid prediction model to estimate an object's future locations based on its pattern information as well as existing motion functions using the object's recent movements. Specifically, an object's trajectory patterns, which have ad-hoc forms for prediction, are discovered and then indexed by their access method for efficient query processing. Yavas et al. (2005) proposed an algorithm to predict the next inter-cell movement of a mobile user in a personal communication systems network. The performance of the proposed algorithm is evaluated through simulations for comparison with two other prediction methods. Sleem and Kumar (2005) proposed a Handoff prediction and enhancement scheme (HOPES) that explores the use of prediction techniques in mobility management to improve the end-to-end traffic quality. The fundamental difference between HOPES and other predictive mobility management techniques is that HOPES uses a topography-aware predictive approach that combines the mobile host's movement history, current state, and the topography of the cells.

Tseng and Lin (2006) proposed a novel data mining method, SMAP-Mine, that can efficiently discover the services mobile users request, associated with their sequential movement patterns. Through empirical evaluation under various simulation conditions, SMAP-Mine is shown to deliver excellent performance in terms of accuracy, execution efficiency and scalability. Tseng et al. (2007) proposed a data mining algorithm named TMSP-Mine to efficiently discover the temporal mobile sequential patterns (TMSPs) of users in LBS environments. This was the first work to focus on mining mobile sequential patterns associated with moving paths and time intervals in LBS environments. They also proposed novel location prediction strategies that use the discovered TMSPs to effectively predict the next movement of mobile users. Lu et al. (2012) proposed a Cluster-based Temporal Mobile Sequential Pattern Mine (CTMSP-Mine) algorithm to discover the Cluster-based Temporal Mobile Sequential Patterns (CTMSPs). They also proposed a prediction strategy that uses the discovered CMSPs to precisely predict the next movement of mobile users.

Lu et al. (2012) proposed a novel framework, called Mobile Commerce Explorer (MCE), to mine and predict mobile users' movements and purchase transactions in mobile commerce contexts. The MCE framework consists of three major components: a Similarity Inference Model, a Personal Mobile Commerce Pattern Mine (PMCP-Mine) algorithm, and a Mobile Commerce Behavior Predictor. Tang, Liao, and Sun (2013) proposed a novel three-stage procedure to elucidate the shopping behavior of mobile users to better predict customer preferences. The framework can discover sequential rules from contextual data and overcome the barrier when multiple dimensions of contextual information are used in the prediction. Ying, Lee, and Tseng (2013) proposed a novel mining-based location prediction approach called Geographic-Tem poral-Semantic-based Location Prediction (GTS-LP), which accounts for a user's geographic-triggered intentions, temporal-triggered intentions, and semantic-triggered intentions to estimate the probability of the user visiting a particular location. The core idea is the discovery of the user's trajectory patterns, namely GTS patterns, to thus capture frequent movements triggered by the three kinds of intentions. Table 1 summarizes the relevant research and compares them with the proposed research in terms of framework structure, mining method, prediction method, behavior context, contexture association, and utility consideration.

3. Research methodology

The proposed shopping behavior prediction (SBP) system consists of four major components: a mobile transaction sequence database, a behavior mining module, a similarity inference module, and a behavior prediction module. The mobile transaction sequence database records the customer's movement and purchasing behavior. The behavior mining module then generates a set of frequent shopping patterns, called high-utility mobile sequential patterns (UMSPs), using the $UMSP_L$ algorithm. In the similarity inference module, the store-to-store (StoS) similarities and item-to-item (ItoI) similarities are derived from the mobile transaction sequence database using the proposed similarity inference algo-