Information Sciences 314 (2015) 28–40

Contents lists available at ScienceDirect

Information Sciences

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

An effective taxi recommender system based on a spatio-temporal factor analysis model

Ren-Hung Hwang, Yu-Ling Hsueh *, Yu-Ting Chen

Dept. of Computer Science & Information Engineering, National Chung Cheng University, Taiwan

ARTICLE INFO

Article history:
Received 16 March 2014
Received in revised form 27 January 2015
Accepted 30 March 2015
Available online 7 April 2015

Keywords:
Data mining
GPS data analysis
Location-based services

ABSTRACT

The taxi fleet management systems based on GPS have become an important tool for taxi businesses. Such systems can be used not only for fleet management, but also to provide useful information for taxi drivers to increase their profits by mining historical GPS trajectories. In this paper, we propose a taxi recommender system for determining the next cruising location, which could be a value-added module in fleet management systems. In the literature, three factors have been considered in different studies to address a similar objective: distance between the current location and the recommended location, waiting time for the next passengers, and expected fare for the trip. In this paper, in addition to these factors, we consider one key factor based on driver experience: what is the most likely location to pick up passengers, given the current passenger drop off location. A location-to-location graph model, referred to as an OFF–ON model, is adopted to capture the relation between the passenger drop-off location and the next passenger get-on location. We also adopt an ON–OFF model to estimate the expected fare for a trip that begins at a recommended location. A real-world dataset from CRAWDAD is used to evaluate the proposed system. A simulator that simulates the cruising behavior of taxies in the dataset and a virtual taxi that cruises based on our recommender system is developed. Our simulation results indicate that although the statistics of the historical data may be different from real-time passenger requests, our recommender system is still effective in terms of recommending more profitable cruising locations.

1. Introduction

Due to the dramatically reduced costs of Global Positioning System (GPS) devices, taxi fleet management systems based on GPS have become very popular for taxi companies. Using such a system, a taxi company is able to monitor the time-stamped GPS trajectories of its taxi cabs. Furthermore, additional information, such as the status of a taxi, including waiting at a stand, cruising, occupied, and off shift, can also be tracked. The GPS taxi fleet management can be used not only for fleet management and security but also to provide useful information for taxi drivers to allow them to earn greater profits by mining historical GPS trajectories and statuses of taxies. As a consequence, many researchers have performed research on methods that can improve the efficiencies of taxi businesses, especially in terms of recommender systems for taxi drivers under different conditions and with different objectives [18,2,17,25,21,24]. For a taxi driver, the issue of most concern is likely to be how to maximize his profit. The daily routine of a taxi driver may consist of several phases of cruising and being occupied.

* Corresponding author.
E-mail addresses: rhhwang@cs.ccu.edu.tw (R.-H. Hwang), hsueh@cs.ccu.edu.tw (Y.-L. Hsueh), wendy10445@gmail.com (Y.-T. Chen).

http://dx.doi.org/10.1016/j.ins.2015.03.068
0020-0255/

Specifically, a taxi driver may cruise the road network while searching for passengers for a short period of time (which may include waiting at taxi stands) and then pick up passengers and drive to the designated destination (occupied time). When the passengers get out of the taxi, the driver starts cruising the road network again. This is when a recommender system could be used to help the driver determine where to cruise such that his profit can be increased. The purpose of this work is to recommend a location for the taxi driver to cruise to so that he can earn a greater profit compared to simply cruising based on his experience.

Several factors shall be considered for guiding a taxi driver to cruise to a more profitable location. First, the distance between the current location and the recommended location should be short to save time and gas [18,2]. Second, when the taxi arrives at the recommended location, the waiting time for the next passenger should also be short [2]. Third, if the taxi driver is able to pick up passengers at the recommended location, a larger fare for the trip is preferred [18]. In the literature, most studies have considered two of these three factors with different approaches toward utilizing the historical data. In this paper, we consider all three of these factors. In addition, we consider a fourth factor by mining the relation between the drop-off location and the get-on location of the next passenger to model the driver's experience, which directly affects revenue. A location-to-location graph, called the OFF–ON model, is proposed. The existing work, which only considers the impact of an individual factor (e.g., vacancy rate or waiting time) may not highly correlate with overall performance in terms of revenue. For example, suggesting a location to a taxi driver solely based on its popularity may not ensure higher revenues in the long term. To solve this problem, we believe that learning the driving patterns of experienced drivers is helpful for enabling increased revenue. Thus, the contribution of this paper is to explore all four factors and analyze the importance of each factor.

A simulator that simulates the cruising behavior of the taxies in the data set and one virtual taxi, which cruises based on our recommender system, is developed. Our simulation results indicate that the revenue earned by the virtual taxi is within the top 1% of all revenues on weekdays and the top 4% of all revenues during weekends based on the use of historical data in the simulation. The results also show that the revenue earned by the virtual taxi is within the top 6% of revenues on weekdays and the top 18% of revenues during weekends when a new data set is used in the simulation. We conclude that although the statistics of the historical data may be different from real-time passenger requests, our recommender system is still effective in terms of recommending a more profitable cruising location. The remainder of this paper is organized as follows. Section 2 discusses the related work. The problem is formulated in Section 3, and our solution is presented in Sections 4–6. A baseline approach is proposed in Section 7. The simulation processes are described in Section 8. Then, Section 9 presents the experiment results. Finally, Section 10 gives the concluding remarks and discusses future research directions.

2. Related work

The field of recommender systems has been intensively studied in past decades. In early studies, a recommendation was made using a collaborative filtering technique (CF for short), which does not consider user preferences and lacks context sensitivity. Adomavicius et al. [1] further described possible extensions by integrating a User-Item matrix that captures the contextual information for use in the recommendation process. In [3], Chu et al. proposed a hybrid system that uses a machine learning model to infer a content-based, personalized recommendation. Javari et al. [10] emphasized the improvement of the performance of the CF algorithm for classic recommender systems and considered a number of elements (e.g., the structure of the user-item bipartite network and user profiles) for reliability extraction. In contrast, Huang et al. [9] emphasized the matrix completion issue that occurs in data mining techniques (e.g., the CF algorithm). Shani et al. [20] discussed various criteria for evaluating the performance of a recommender system. Due to the proliferation of social network applications, the check-in data, which enable a user to share their experiences, have become widely available. Therefore, in [13], the HITS-based model was deployed to rank popular venues and experienced users by analyzing the check-in data to form a popular user-venue matrix for use by the recommender system. Long et al. [14] focused on designing a POI recommendation algorithm based on the social relations to provide POI recommendations. Unlike most studies that do not consider both the diversity and precision of recommendations, Javari et al. [11] built a probabilistic model by learning from historical data sets and used the model to generate recommendations.

A taxi recommender system can be thought of as an extension of recommender systems to different objectives. We have emphasized this particular application. Next, we describe the related work. Chang et al. [2] proposed an on-line prediction application to demonstrate a context-aware taxi demand system using a real maps. They categorized the landmarks or the classes of roads using a location ontology and performed context-aware pattern mining on taxi request records by adopting different clustering approaches. Customer demand can be understood through these processes. However, their approach did not validate the demand of the location in different situations in practice. In our work, we find that taxi demands vary in different time slots, and we validate the effect using statistical utilities. Moreira-Matias et al. [17] focused on recommendations for the next taxi stand. They adopted time-series forecasting techniques to predict the spatio-temporal distribution in real time. Their results showed that the waiting time to pick-up a passenger could be reduced by 5%. Powell et al. [18] proposed the Spatio-Temporal Profitability (STP) map for guiding taxicabs to cruise to more profitable locations. They mainly focused on modeling the profitability of locations. To reduce the travel time to the recommended location, the STP map only considers a sub-region around the taxi cab's current location, which is further divided into a grid of equally sized cells. The delay between the current location and the recommended cruising location was also considered. Takayama et al. [21]