From Data to Knowledge to Action :
A Taxi Business Intelligence System

Chenggang Wang
Southwest China Institute of Electronic Technology
Chengdu, China.
cgwang@126.com

Huaixin Chen
Southwest China Institute of Electronic Technology
Chengdu, China.
Chenhuaixin@sina.com

Wee Keong Ng
School of Computer Engineering
Nanyang Technological University
Singapore
wkn@acm.org

Abstract— **Taxis play an important role in offering comfortable and flexible service within Singapore's public transport system. Due to the inherent randomness in taxi service system, many taxi companies still rely on the drivers' experience to seek passengers. Today, Singapore's five taxi companies now use some form of wireless and GPS (Global Position System) satellite to track taxis traveling in urban area. GPS-equipped taxis can be viewed as ubiquitous mobile sensors which enable us to collect large amounts of location traces of individuals or objects. In this paper, we first investigate the characteristics of travel behavior of urban population. Next, a taxi business intelligence system is proposed to explore the massive transportation data based on spatial-temporal data mining techniques. Furthermore, various taxi business models are created to make comprehensive analysis on taxi business problems. Finally, the value of the taxi business intelligence system is demonstrated by applying it to some real-world scenarios. Results show that this system can significantly improve the quality of taxi services.**

Keywords-inefficient taxi business;GPS equipment; spatial-temporal data mining; decision making; intelligent system

I. INTRODUCTION

Taxis play an important role in offering a comfortable and flexible service within Singapore's public transport system. However, customers and taxi drivers sometimes experience frustration while seeking taxis and passengers respectively. For example, taxis may be waiting at a vacant stand while customers may be queuing in vain elsewhere. This problem has baffled taxi service ever since it existed [1].

Today, Singapore's taxis have been equipped with GPS receiver and some form of wireless communication device, in order to report its location to taxi monitoring control center [2]. The in-vehicle telematics device builds a location record, including vehicle ID, timestamp, speed, operation status, longitude and latitude. The status field indicates the current operation information of the taxi, specifying whether the taxi is carrying a passenger or empty. In the taxi monitoring control center, location records are not discarded but are being accumulated, since those data have much information on the movement history of each vehicle, taxi service time, dispatch performance, and so on [3]. Accordingly, it provide us with an unprecedented opportunity to automatically extract useful knowledge, which in turn deliver intelligence for real-time decision making in various fields, such as location recommendations, online taxi booking services and taxi business management.

In this paper, the characteristics of travel behavior of urban population are first analyzed to build a spatial-temporal description of potential customers. Second, a framework of intelligent taxi business system is proposed to explore the large-scale transportation datasets, and to extract the value-added information. Definitely, it can be used as guidance for reducing inefficiencies in fuel consumption of transportation sectors, improving customer satisfaction, and business performance. For example, "we can estimate the waiting time distribution and pick-up location pattern in a specific place at a specific time instant, hence systematically recommends the waiting spot to the driver of a currently empty taxi" [4]. Finally, the application examples are designed to show how to improve customers' satisfaction and drivers' performance by using taxi business models and real-time traffic data.

The rest of this paper is organized as follows. The characteristics of urban population travel behavior are analyzed in SectionII; the component of taxi business intelligence system is introduced and discussed in Section III; in Section IV, we demonstrate the value of our system by applying it to some real-world scenarios; finally, we conclude the paper in SectionV.

II. TRAVEL BEHAVIOR CHARACTERISTICS

A. Spatial Characteristics

Diverse activities in urban area led to the development of different functional space within the urban area. Urban population lives and works in these spaces, whose spatial characteristic can be expressed by its spatial location. Daily travel activities always happen between different functional spaces. To some degree, urban functional space distribution determines the spatial characteristics of urban population travel behavior characteristics [6].

Singapore is an island city-state in Southeast Asia. About 5 million people live and work within 700 square kilometers. The city can be divided into five regions, as shown in Fig. 1[7].

Figure 1. Regions of Singapore

- Central region contains most of the high-rise office buildings and main commercial centre in the city, such as Orchard Road, Singapore's main retail shopping district. Generally, there has a high density of taxi passengers, as well as a high density of taxis [2].

- Other four regions include most of Singapore's public housing and industrial estates. These four regions have a low density of taxi passengers and a low density of taxis as well [7][8].

B. Temporal Characteristics

Taxi demands vary dynamically over time, as shown in Fig. 2 that indicates the pick-up frequency distribution of one thousand taxis which were selected at random. The demand histogram typically have a pronounced peak in office hour start time, and a more elongated peak in office hour end time, with demand slowly tapering out into the daytime, night, and the dawn.
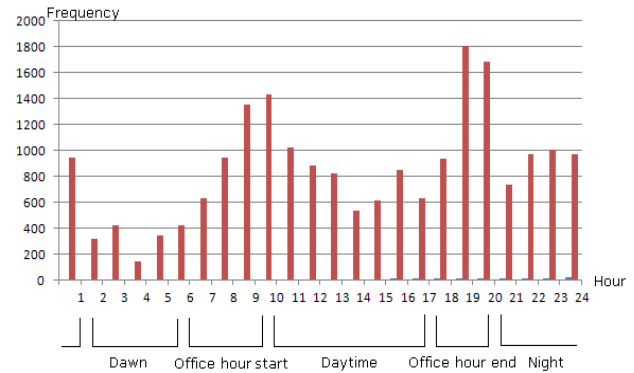

Figure 2. Daily hour clustering

In addition, in different regions the temporal characteristics of taxi demand are very different. For example, city hall has many pick-ups especially in the night time, namely, from 9 PM to 5 AM of next day. During this period, there is no public bus transportation, but city hall area still gathers many people. Oppositely, airport area has high pick-up ratio during the airplane arrival and departure time [2][20].

III. SYSTEM ARCHITECTURE

Extracting value-added information from large-scale transportation datasets is not a trivial task. Conventional data analytic tools are usually not customized for handling large, complex, dynamic, and distributed nature of traffic information. Thus, an integrated information processing and analyzing platform has been designed to help taxi companies to improve their business performances by understanding the behaviors of both drivers and customers. These procedures can be divided into three layers: data layer, pattern discovery and modeling layer, and decision layer, as shown in Fig. 3[9][10].
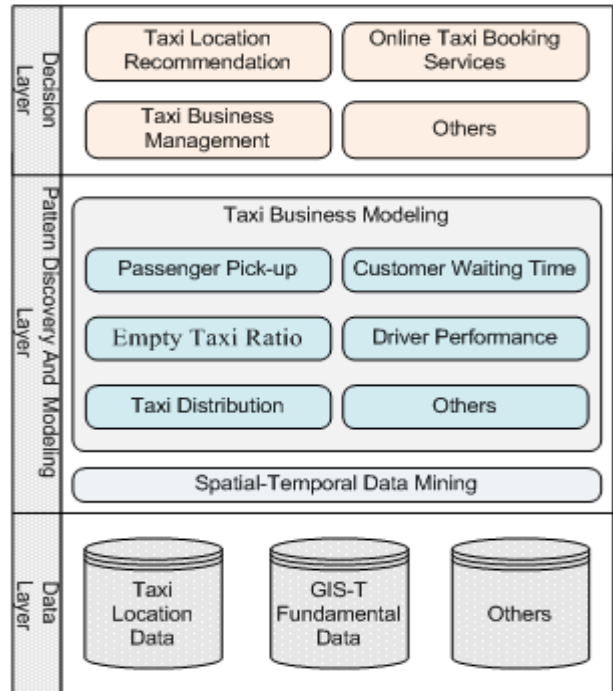

Figure 3. System Overview

A. Data Layer

- Taxi location data is automatically collected by telematics systems equipped by taxis. Raw taxi locations are filtered, classified and integrated into some form of database, and then they can be conveniently accessed by researchers.

- GIS-T (Geography Information System - Transportation) [12] fundamental data provides all of the road information, land use information, traffic zone information, and facility information. Through GIS-T fundamental data, the spatial analysis is done by converting latitude and longitude to street address and modeling mobility patterns of taxis.

- Other data sources include passenger survey results, taxi driver survey results, etc. Based on this information, we evaluate our system and adjust the operational settings.

B. Pattern Discovery And Modeling Layer

- Spatial-Temporal data mining is the core function of pattern discovery and modeling layer, it can offer different types of analysis models for different purposes, processing, analyzing and gaining various thematic data. The spatial-temporal data mining process usually consists of three phases: (1) pre-processing or data preparation; (2) modeling and validation; and (3) post-processing or deployment [13]. During the first phase, the raw data collected from ubiquitous mobile sensors is cleaned and organized for the mining stage. Some of data received from taxis contains noise and missing values caused by malfunction of the GPS unit or transmission problems, a data cleaning must be performed to identify and remove these data to assure the data quality. On the other hand, because the expected traffic daily models are different in different regions. Public holidays also have unique traffic patterns. A categorization is needed to separate the traffic data into different datasets. The second phase consists of association analysis, cluster analysis and data classifications, creating different analytical patterns that accurately describe the taxi business. Widely-used algorithms are available in software packages like WEKA [10], also being used for this research. For example, k-means algorithm can be used to group spatial-temporally similar pick-up and drop-off points as interests to these areas vary throughout time of the day, day of the week, and even seasons of the year. The third phase consists of using the model, evaluated and validated in the second phase, to effectively study the application behavior [14][15].

- Passenger pick-up pattern classification is one of the most basic and useful analyses for taxi business. The analysis begins with the extraction of pick-up record from the raw transportation data. "This step inevitably has to handle a large amount of data from the widely distributed area over a long time interval. Then, spatial-temporal grouping is needed, as the different area has the different traffic pattern due to the uniqueness of facilities and road segment distribution. Within an area, it is desirable to make a further refined cluster to be able to design a taxi location recommendation system" [3][16].

- Customer waiting time model and empty taxi ratio model are very important indicators of evaluating taxi service quality, service capabilities and customer satisfaction. From a taxi driver's point of view, an efficient taxi system is one in which his or her taxi is never empty. By contrast, a customer's idea of an efficient taxi system is one that delivers an empty taxi to his or her location at the instant one is desired. Obviously these goals are in conflict [17]. Taxi operators need to study the relationship between taxi demand and supply, and balance the interests of these two stakeholder groups while running an efficient and profitable taxi business.

- Driver performance model is the model that represents the indicators of taxi drivers' service capability and work efficiency. This model is related to driving hours, occupancy rates and fee achievement. Based on this model, successful taxi drivers can be separated from other drivers, and efficient business knowledge can be extracted from the historical data of successful taxi drivers for helping other taxi drivers to improve their business performances.

- Real-time taxi distribution information is usually regarded as an indicator of taxi supply ability in different regions. To describe taxi distribution, the urban area has been divided into some grids. Through statistical analysis on taxi spatial density in each grid, the real-time distribution of empty taxis can be obtained. This information will play an important role in taxi location recommendation and taxi business management [18][19].

C. Decision Layer

- Taxi location recommendation function is very important for the improvement of taxi drivers' performance. After estimating the waiting time distribution and pick-up location pattern, this recommender provides empty taxi drivers with some locations, towards which they are more likely to pick up passengers quickly and minimize the empty taxi ratio [21][22][23].

- Online taxi booking service function is different from most of the current taxi booking system. This service can be performed by customer's mobile devices with GPS capability. Once a customer starts a taxi booking process, his or her accurate location will be transmitted to the taxi monitoring control center. Then a taxi with the shortest-time path can be found and dispatched by this system. Online taxi booking system minimize the misunderstanding between taxi drivers and customers, and improve customer satisfaction by reducing taxi waiting time and visualizing the approaching taxi on map.

- Efficient taxi business management is essential for the provision of quality customer service and achievement of high business profit. Taxi business intelligence system can produce various types of taxi business analyses results, which help the manager to make right decisions.

IV. APPLICATION EXAMPLES

The stakeholders of the proposed solution are primarily taxi drivers, customers, and taxi operators. Three examples are designed to demonstrate the value of the taxi business intelligence system by applying it to some real-world scenarios [24][25][26][27][28].

A. Example 1: Looking for customers

Alex is either a new taxi driver or someone who has been in this line of work for years. Alex has just completed a trip and the taxi is currently empty. Alex sends a request to the taxi business intelligence system to provide a sense of where he could go. The taxi business intelligence system starts the location recommendation function immediately. This procedure is described by Fig. 4.
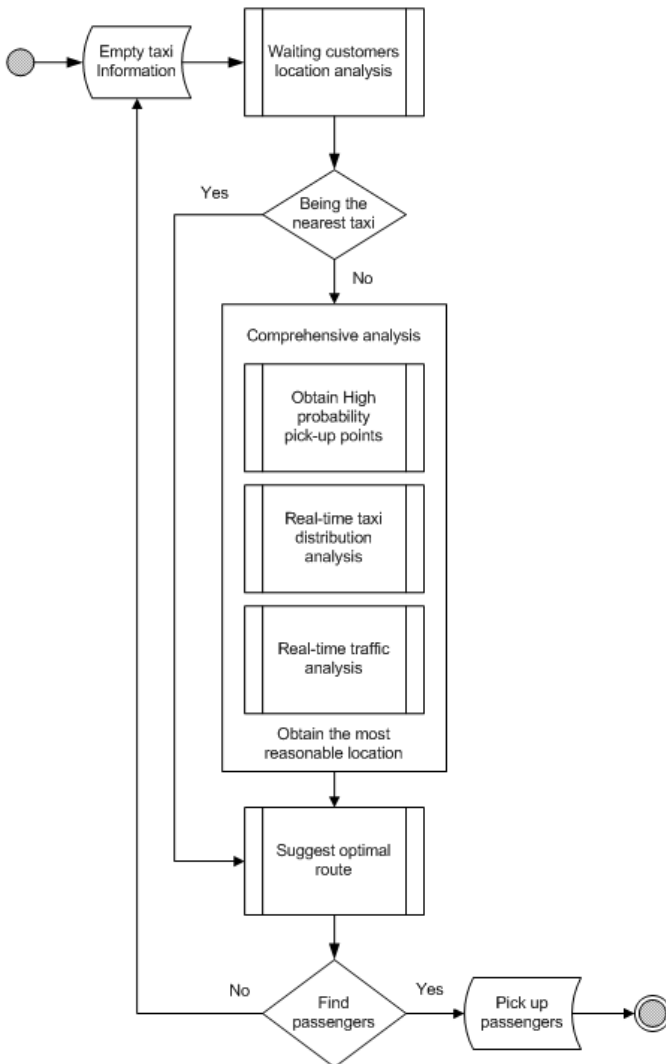

Figure 4. Location Recommendation Procedure

First, the location information of the empty taxi must be analyzed based on waiting customers' location information. If the system find that this taxi is the nearest one to a customer, then it informs Alex that there is currently a customer waiting at location X for a taxi and asks if Alex is committed to pick up this customer. If so, the customer is informed of the taxi number. If there is no customer in the vicinity, the system informs Alex of (one-three) possible locations with various likelihoods of finding customers based on spatial-temporal analytics results performed in the background. In addition, Alex can access location-based mapping and routing, which results in the minimal driving distance before the pick-up event. With the help of the location recommendation system, Alex is now able to make a more informed decision as to where to go next, and his income is increasing at the same time [29].

B. Example 2: Looking for a taxi

Cindy likes to take a taxi home after shopping. She needs a taxi now. She uses her mobile device to connect to online taxi booking services. Then, the system looks for a committed taxi driver and informs Cindy of the taxi number and estimated waiting time. Cindy is able to use her mobile device to visualize and track the route of the impending taxi approaching her, as shown in Fig. 5.
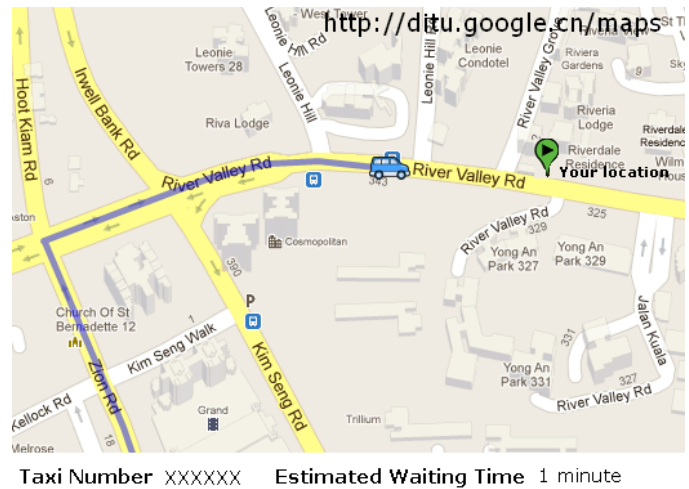

Figure 5. Demo of Online Taxi Booking

This function minimizes the misunderstanding between taxi drivers and customers, and improves customer satisfaction by reducing taxi waiting time and visualizing the impending taxi approaching on map.

C. Example 3: Taxi Business Management

The taxi business intelligence system provides a collection of analytical models and thematic information which a taxi company manager may invoke to make all kinds of decisions, such as taxi requirements measurement, high performance taxi driver selection and real-time taxi business monitor [32][33].

- Accurately taxi requirements measurement is very important for perfectly balancing taxi supplies and demands. The taxi business intelligence system can provide statistical analysis results on customers' waiting time and empty taxi ratio. With this

information support, the manager can forecast the additional taxi demands.

- High performance taxi drivers are the most precious wealth of taxi companies, who typically have higher customer occupancy rates and lower fuel consumption rates. The taxi business intelligence system can help the manager select high performance drivers and extracting successful business knowledge for helping other taxi drivers to improve their business performances.

- Real-time taxi business monitoring is essential for the provision of quality taxi services. It can help the manager to deal with all kinds of emergencies by adjusting the operational settings of the taxi business intelligence system. The most typical emergency is unexpected huge amount of taxi demands after a large-scale concert.
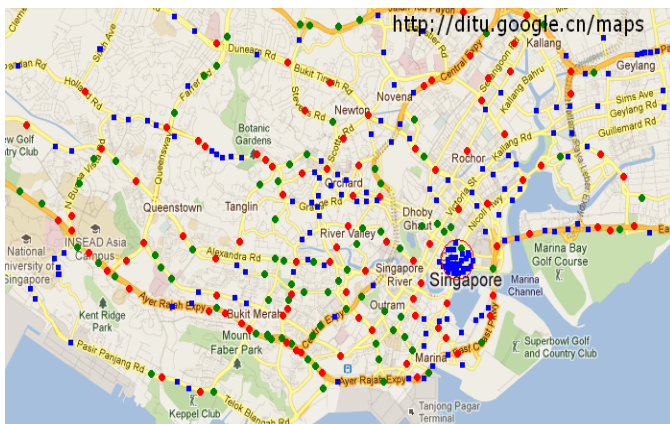

Figure 6.   Demo of Real-time Taxi Business Monitor

Fig.6 shows that an unexpected huge amount of taxi demands generate near city hall. The manager finds out the emergency from the monitor interface, and changes pick-up pattern by adjusting the grade of city hall location cluster. Then, the location recommendation system informs the nearby empty taxis as soon as possible.

V.   CONCLUSION

Because of the inherent randomness in taxi service system, the conventional taxi industry is low efficiency, high fuel consumption and low customer satisfaction. This paper has built a taxi business intelligence system for Singapore's taxi industry, through which large-scale transportation datasets can be analyzed, and value-added information can be extracted based on spatial-temporal data mining technologies. At the same time, this paper analyzed the characters of population travel behavior, and then put forward various taxi business models to describe both passengers and taxi drivers' behavior. As a result, the taxi business intelligence system can overcomes many of the limitations for existing taxi business system. This system can be effectively used in taxi business to improve the quality of taxi services.

As future work, we are planning to obtain more taxi transportation data and mine more taxi business knowledge from it. In addition, in order to adapt to the cloud computing trend, this system will be performed on cloud platform.

REFERENCES

[1] Q. B. Meng, S. Mabu, L. Yu, and K. Hirasawa, "A Novel Taxi Dispatch System Integrating a Multi-Customer Strategy and Genetic Network Programming," Journal of Advanced Computational Intelligence and Intelligent Informatics, Vol.14, pp. 442-452 ,Jul 2010.

[2] D. Santani, R. K. Balan and C. J. Woodard, "Spatio-Temporal Efficiency in a Taxi Dispatch System," Singapore Management University, Singapore ,Oct 2007 .

[3] J. Lee, I. Shin and G. L. Park, "Analysis of the Passenger Pick-Up Pattern for Taxi Location Recommendation," Fourth International Conference on Networked Computing and Advanced Information Management, Vol.1, Gyeongju, Korea, Sep 2008, pp.199-204.

[4] J. Lee, "Analysis on the Waiting Time of Empty Taxis for the Taxi Telematics System," Third International Conference on Convergence and Hybrid Information Technology,Vol.1, Nov 2008, pp.66-69.

[5] T. Takayama, K. Matsumoto, A. Kumagai, N. Sato, and Y. Murata, "Waiting/Cruising Location Recommendation for Efficient Taxi Business," Internation Journal of Systems Applications, Engineering & Development, Vol.5, no.2, pp.224-236, 2011.

[6] Y. Zheng, Y. Liu, J. Yuan and X. Xie, "Urban Computing with Taxicabs," UbiComp'11, Beijing , China, Sep 2011, pp.17-21.

[7] http://en.wikipedia.org/wiki/Regions_of_Singapore.

[8] Department of Statistics Singapore. Statistics Singapore-Census of Population 2010, Households and Housing, Feb.2011. http://www.singstat.gov.sg/pubn/popn/c2010sr2.html.

[9] K. T. Seow, N. H. Dang and D. Lee, "A Collaborative Multiagent Taxi-Dispatch System," IEEE Transactions on Automation Science and Engineering , Vol.7, no.3, pp. 607-616, 2010.

[10] R. Geetha, N. Sumathi and Dr. S. Sathiyabama, "A Survey of Spatial ,Temporal and Spatio-Temporal Data Mining,", Journal of Computer Applications, Vol.1, no.4, pp.21-23, 2008.

[11] http://www.cs.waikato.ac.nz/ml/weka/.

[12] Y. FU, Z. LI, K. SONG, Z.QIU and X. MA, "Integrated Traffic Management Platform Design Based on GIS-T," 6th International Conference on ITS Telecommunications Proceedings, Chengdu, China, Jun 2006, pp.29-32.

[13] H. Zhu, Y. Zhu, M. Li and L. M. Ni, "SEER: Metropolitan-scale Traffic Perception Based on Lossy Sensory Data," IEEE INFOCOM, Brazil, pp. 217-225, Apr 2009.

[14] J. Yuan1, Y. Zheng, L. Zhang, X. Xie and G. Sun, "Where to Find My Next Passenger?," UbiComp'11, Beijing, China, pp.17–21, Sep 2011.

[15] Y. Zhao, "Mobile Phone Location Determination and Its Impact on Intelligent Transportation Systems," IEEE Transactions on Intelligent Transportation Systems, Vol. 1, no. 1, pp. 55-64, Mar 2000.

[16] J. Mennis and J.W. Liu, "Mining Association Rules in Spatio-Temporal Data," Department of Geography, University of Colorado, USA.

[17] D. Santani, R. K. Balan ,C. J. Woodard, "Understanding and Improving a GPS-based Taxi System," 6th International Conference on Mobile Systems, Applications, and Services, USA, Jun 2008.

[18] J. F. Ehmke, D. C. Mattfeld, "Data Allocation and Application for Time-Dependent Vehicle Routing in City Logistics," European Transport – International Journal of Transport Economics, Engineering and Law, no. 46, pp.1-12, 2010.

[19] J. Hu, W. Cao, J. Luo, X. Yu, "Dynamic Modeling of Urban Population Travel Behavior based on Data Fusion of Mobile Phone Positioning

Data and FCD," Proceeding of the 17th international conference on Geoinformatics, Bethesda, USA, Aug 2009, pp.12-14.

[20] Ashish, Agarwal, "A Comparison of Weekend and Weekday Travel Behavior Characteristics in Urban Areas," University of South Florida, 2004.

[21] Y. J., J. Dai, C. T. Lu, "Spatial-Temporal Data Mining in Traffic Incident Detection," In :SIAM Conference on Data Mining, USA, Apr 2006.

[22] G. Chen, X. Jin, J. Yang, "Study on Spatial and Temporal Mobility Pattern of Urban Taxi Services," The Proceedings of ISKE, Hangzhou, China, Nov 2010, pp. 422-425.

[23] S. F. Cheng, X. Qu, "A Service Choice Model for Optimizing Taxi Service Delivery," 12th International IEEE Conference on Intelligent Transportation Systems, ITSC, USA, Oct 2009, pp.1-6.

[24] K. Zhang, H. Li, "Fusion-Based Recommender System," the 13th international conference on information Fusion, Edinburgh, UK, Jul 2010, pp.1-7.

[25] P. Haluzová, "Effective Data Mining for a Transportation Information System," Acta Polytechnica Vol. 48 No. 1, pp24-29, 2008.

[26] Y. Ge, C. Liu, H. Xiong, J. Chen, "A Taxi Business Intelligence System," KDD'11, San Diego, USA, pp.735-738, Aug 2011.

[27] Z. Liao, "Real-time taxi dispatching using global positioning systems," Communication of the ACM, Vol. 46, pp. 81-83, 2003.

[28] J. Lee, "Traveling Pattern Analysis for the Design of Location-Dependent Contents Based on the Taxi Telematics System," International Conference on Multimedia, Information Technology and its Applications, Jul. 2008.

[29] S. Shekhar, "Transportation Data Mining Challenges," University of Minnesota, www.cs.umn.edu/~shekhar.

[30] G. Andrienko, D. Malerba, M. May and M. Teisseire, "Mining Spatio-Temporal Data," J Intell Inf Syst , Vol 27, pp.187-190, 2006.

[31] J. S.Taylor, T. D. Bie, N. Cristianini, "Data Mining, Data Fusion and Information Management," IEE Proceedings - Intelligent Transport Systems Vol.153, no.3, pp. 221-229, Sep 2006.

[32] R. K. Balan, N. X. Khoa, and L. X. Jiang, "Real-Time Trip Information Service for a Large Taxi Fleet," MobiSys'11, Bethesda, USA, pp. 99-112, Jun 2011.

[33] H. Kim, J. S. Oh, and R. Jayakrishnan, "Effect of Taxi Information System on Efficiency and Quality of Taxi Services," Journal of the Transportation Research Board, no.1903, pp. 96–104, 2005.