論文の内容の要旨
Thesis Summary

論文題目
TAXI GPS DATA ANALYSIS FOR THE IMPROVEMENT OF TAXI OPERATION OF BANGKOK
(バンコクにおけるタクシー運行改善のための GPS データ分析)

氏 名
Saurav Ranjit (ﾗﾝｼﾞｯﾄ　ｿｳﾛﾌﾞ)

本文

With the growing advancement in the field of Global Positioning System (GPS) technology, the utilization of GPS in the field of science has increased significantly in the past recent years. One of the prominent fields which have benefited from the GPS technology is the spatial information science. Spatial information is essentially a digital data that provides information concerning location, people and in times their activities as well. In such, the transportation sector is one of the most benefitted sectors with GPS embedded technology. Whether it is navigation of a vehicle or tracking, GPS technology has been the front-runner for providing the information which further helps decision making. However, technology is not just limited to navigation or tracking. In a recent year, many big cities like New-York and Beijing have started embedding GPS device in the taxi vehicle to collect traffic information. Such vehicle is essentially known as floating car or a probe car. Taxi service are ubiquitous all over the world as a convenient way of commuting in big cities. Bangkok, capital of Thailand, is no exception as more than 100,000 taxis approximately runs daily in and around the city. As, taxi are operational throughout city, mobility data from these vehicles can be an asset for governing urban management and planning. Acknowledging the fact, Toyota Tsusho Nexty Electronics (Thailand) Co. Ltd (TTNET) Bangkok, Thailand has equipped approximately 10,000 GPS devices onto the taxi running in Bangkok city and surrounding provinces. GPS device equipped onto the taxis collects spatial-temporal information every 3-5 seconds with approximately 50 million GPS points per day. Speed, direction and taxi meter status are also collected. Due to the accuracy of GPS, collected data is not always precise and are not always on the road segment. Primarily, GPS data from these probe taxies are utilized for providing the traffic information on a major road segment, however utilization of probe data is not limited to it. The existing literature does shows that there are issues related with the taxi operation in Bangkok, Thailand whether it is from the driver perspective or it is from the passenger perspective. Spatial and temporal data are available from the taxi operation from the Bangkok and surrounding region. These data could be a value asset which could help improve the operation of taxi service through data mining technology. However, lack of proper data infrastructure management system could be the hindrance if proper and efficient mining technology needs to be applied. The detail issue as related is presented as following.

Proper Data Infrastructure Management System: Without proper data infrastructure management system, the data mining working could be a very challenge task especially when dealing with the big data volume. Spatial data involving mobility data from the vehicle movement are constantly

increasing. In such cases, how to properly handle the data becomes the primary task before any other data mining algorithm could be any applied.

Taxi Operation Modeling with Quantitative Data Evidence: The issue with the taxi operation services are exist as shown from the past literature. However, model to understand the behavior based on quantitative data evidence are not properly established yet. If the proper behavior model is not established it could pose a challenge when dealing with the ways to improve the service level of the taxi operations.

Taxi Optimization Modeling: The main two fundamental objectives of the taxi business or the service is to provide good service to the customer or passenger and in turn obtain the monetary profit. However, from the data evidence it is clear that there are issues related with both providing good service as well as getting better monetary profit. Taxi passenger are not happy when they are not provided with the good taxi service or when they are rejected to the service itself. On the other hand, taxi drivers are not getting enough passenger. Though the situation is ironic in nature itself, the problem does exist. One of way to minimize the issue is to optimize the operation of the taxi service. The optimization method as proposed will have ability for the driver in which driver can choose passenger depending upon the passenger origin and destination as well as available demand in the region. Optimization model is to provide recommendation to taxi driver which passenger would be better to choose and which not through mobile application. The hypothesis behind the model is that when driver have ability to choose the passenger then passenger rejection would be drastically minimized as well as choosing passenger would give driver some degree of freedom on how monetary profit could be improved. In addition, optimizing route for efficient taxi operation also plays an important role determining how much profit the driver can make by reducing the operation cost on fuel as well as its maintenance.

The main objective of this research is to help improve taxi operation in Bangkok region through quantitative data analysis from the GPS probe data from taxi. The overall objective is subclass as following

- Develop the data infrastructure management system for big mobility data handling and operation.
- Develop the taxi simulation model of the taxi operation for the Bangkok and the surrounding region
- Develop the optimization model for the improvement of the taxi operation

The objective is designed to address the issue for taxi operation in Bangkok which would provide the taxi driver with assistance that would increase the income, reduce the working hour in turn provide better service level for the passengers.

The research task starts with the development of the data infrastructure. The data infrastructure is further categorized for handling probe data and road network data. Bangkok taxi survey is also conducted for the data preparation. Finally, a data platform namely Horton Data platform is developed to handle the big spatial dataset. The second research task includes taxi simulation model. The simulation model is constructed with multiple variables that are derived from the GPS probe taxi data. The third and the final task includes the optimization of the taxi operation based

on proving assistance to the taxi driver to improve the overall taxi operation in the Bangkok and the surrounding regions.

The first part of the research is dedicated to analyze various map matching techniques that can be utilized efficiently and accurately for big GPS dataset. Road network for map-matching is obtained from the Open Street Map (OSM). OSM road network is cleaned for topological error like floating links, pseudo nodes using spatial operation on PostgreSQL. Total of 1,107,798 road link feature is extracted for Thailand which is converted to Well Known Text (WKT) for using in map matching process. Map matching is performed considering road geometry, topology, and connectivity. The simplest geometry operation for matching operation is buffer operation which finds line segment from GPS point within buffer distance. However, identifying optimum buffer distance is difficult as point accuracy and complexity of the road link is different at various locations. Hausdorff distance matrix is applied for topological operation by taking consecutive 4 GPS points. The Hausdorff distance matrix measures degree of similarity between road link geometry with the line segment geometry created from 4 consecutive GPS point. Topology operation improved the accuracy, but its accuracy could be compromised near the road intersection where consecutive GPS point lies between different road segments. Finally, probabilistic approach in which both geometry, topology is considered along with new parameter of road link connectivity. The probabilistic approach is performed using Hidden Markov Model with Hausdorff distance matrix. Distance matrix provided road link candidate of GPS point for which initial probability, measurement probability and forward transition probability is measured, and GPS point is matched with road link. Accuracy assessment for all the map matching technique is performed for various cases such as 'single and multiple lane road', 'road intersection' etc. As mentioned, both data as well as road link are very huge, computation time is a critical factor for efficient every operation. A distributed computing platform for large scale data which utilized Hadoop/Hive is used for all computation involved in map-matching operation. The result obtained is labelled GPS data, based on road network, which is utilized for accurately map traffic congestion, improve taxi operations through optimizing taxi routes. By using OSM road networks, the techniques could be replicated and implemented for other country where OSM is available.

The second part of the research is dedicated to model the taxi operation in the Bangkok. Taxi behavior is a spatial–temporal dynamic process involving discrete time dependent events, such as customer pick-up, customer drop-off, cruising, and parking. Simulation models, which are a simplification of a real-world system, can help understand the effects of change of such dynamic behavior. An agent-based modeling and simulation is utilized, that describes the dynamic action of an agent, i.e., taxi, governed by behavior rules and properties, which emulate the taxi behavior. Taxi behavior simulations are fundamentally done for optimizing the service level for both taxi drivers as well as passengers. Moreover, simulation techniques, as such, could be applied to another field of application as well, where obtaining real raw data are somewhat difficult due to privacy issues, such as human mobility data or call detail record data. This research describes the development of an agent-based simulation model which is based on multiple input parameters (taxi stay point cluster; trip information (origin and destination); taxi demand information; free taxi movement; and network travel time) that were derived from taxi probe GPS data. As such, agent's parameters were mapped into grid network, and the road network, for which the grid network was used as a base for query/search/retrieval of taxi agent's parameters, while the actual movement of taxi agents was on the road network with routing and interpolation. The results obtained from the

simulated taxi agent data and real taxi data showed a significant level of similarity of different taxi behavior, such as trip generation; trip time; trip distance as well as trip occupancy, based on its distribution. As for efficient data handling, a distributed computing platform for large-scale data was used for extracting taxi agent parameter from the probe data by utilizing both spatial and non-spatial indexing technique.

The last and final part of the research work focuses on the optimization of the taxi behavior model. For the case of optimization scenario is considered i.e. how the driver pick up the passenger. In this scenario, driver have the ability to choose which passenger to pick and which not to pick. The reason behind of having driver choose the passenger is that the refusal rate would decrease drastically. This does not guarantee that all taxi driver is willing to server the customer. However, if certain number of taxi are not willing to server the customer, there will exist other group of vacant taxi driver that are willing to serve the customer making the demand and supply theoretically in the state of equilibrium unless demand out run the supply.