Waiting/Cruising Location Recommendation Based on Mining on Occupied Taxi Data

TSUYOSHI TAKAYAMA1, KENJI MATSUMOTO2, AYAKA KUMAGAI3, NOBUYOSHI SATO1, and YOSHITOSHI MURATA1
1 Graduate School of Software and Information Science
Iwate Prefectural University
152-52, Sugo, Takizawa, Iwate020-0193
JAPAN
{takayama, nobu-s, y-murata}@iwate-pu.ac.jp http://www.iwate-pu.ac.jp/

2 Embedded Business Division
IX Knowledge Inc.
3-22-23, Kaigan, Minato, Tokyo108-0022
JAPAN
subnoad@gmail.com http://www.ikic.co.jp/

3 Faculty of Software and Information Science
Iwate Prefectural University
152-52, Sugo, Takizawa, Iwate020-0193
JAPAN
g031f054@s.iwate-pu.ac.jp http://www.iwate-pu.ac.jp/

Abstract: - Recently, lots of researchers are attracted to constructing information system for efficient taxi business. In general, there are three types of methods in order to catch a passenger for a taxi driver: ‘waiting’, ‘cruising’, and ‘wireless/order’. Conventional systems decline to support ‘wireless/order’ method, and it is not sufficient to support ‘waiting’ and/or ‘cruising’ one. Therefore, in the present paper, we try to support ‘waiting’ and ‘cruising’ based on mining of occupied taxi data and try to catch a passenger more efficiently. According to the result of evaluation experiment, our proposition is effective.

Key-Words: - ITS(Intelligent Transport Systems), taxi, data mining, recommendation, and database.

1 Introduction
Recently, lots of researchers are attracted to ITS (Intelligent Transport Systems) area[1][2][3][4]. Among such a trend, some information systems for efficient taxi business are being constructed. In general, there are three types of methods in order to catch a passenger for a taxi driver: ‘waiting’, ‘cruising’, and ‘wireless/order’. In ‘waiting’, the driver waits a passenger with stopping their taxi in a location. In ‘cruising’, the driver makes their taxi run with looking for a passenger. ‘Wireless/order’ is the method which the driver receives a reservation through taxi operator and goes to the passenger. Automatic Vehicle Monitoring (AVM) system is representative as a conventional taxi information system and mainly supports ‘wireless/order’ method. However, we can not always say that it is sufficient to support ‘waiting’ and/or ‘cruising’ method. Therefore, in the present paper, we execute waiting/cruising location recommendation based on mining of occupied taxi data and try to catch a passenger more efficiently and enlarge resulting income.

2 Previous Studies
In the paper[5], Liao surveys the method of AVM system collaborating with GPS (Global Positioning System) receiver. It is used in ‘wireless/order’ method. Hariharan et al. try to represent and store trajectory from continuous tracking of location, not restricting a taxi, for a moving object[6]. They do not take into account the status of on/off such as in a taxi. In the paper[7], Silva proposes to send a taxi based on the location information of passenger from their GPS terminal. It is also for ‘wireless/order’ method. It does not forecast a passenger’s demand and not
recommend a location to taxi drivers. Hou try to connect a taxi’s supply chain as efficiently as possible[8]. In the paper [9], Lee et al. analyze passenger pick-up pattern based on the location history data collected from a telematics system, which is one of the information provision services for a future car, using mobile communication and internet such as AVM system. They applies K-means method[10] in order to cluster each trajectory data of occupied taxi. They also analyze them in time series.

These researches do not concretely recommend a location where a taxi should waits or cruises.

3 Waiting/Cruising Location Recommendation Based on Mining of Occupied Taxi Data
In this section, we describe our four experiments and their tries.

3.1 Common Matters through Four Experiments

3.1.1 Analysis method
We analyze the occupied taxi data which each taxi driver records in the form of paper media, and find trend from them.

3.1.2 Recommendation method
We provide each taxi driver with location information where we recommend as waiting/cruising location using paper media we call ‘information provision sheet’. Concretely, we provide them with pick-up numbers ranking per location.

3.1.3 Evaluation method
We carry out the following three questionnaire surveys after each experiment:
(i) five levels subjective evaluation that the corresponding information provision sheet has been useful or not.
(ii) either they have changed their waiting/cruising location or not, based on the information provision sheet.
(iii) mean time from start of waiting/cruising to picking-up a passenger
From (ii) and (iii), we analyze a difference based on change of waiting/cruising location or not according to our recommendation.

3.1.4 Collaboration taxi company
We ask to provide us with occupied taxi data and to collaborate with our experiment to ‘H’ company in Morioka-City, which is the capital city of Iwate prefecture in Japan. It has 150 taxi drivers.

3.2 1st Experiment: Investigation of Time Granularity

3.2.1 Method

Table 1 Period of 1st Experiment

<table>
<thead>
<tr>
<th>experiment period</th>
<th>one week from 3th Friday in July</th>
</tr>
</thead>
<tbody>
<tr>
<td>analyzed past data</td>
<td>from Jul. 18th (Fri.) to 24th(Thu.), 2008</td>
</tr>
<tr>
<td>recommendation execution</td>
<td>from Jul. 17th (Fri.) to 23th(Thu.), 2009</td>
</tr>
</tbody>
</table>

Fig.1 Three Alternatives of Time Granularity.

We make picking-up numbers Top-3 ranking at the three types of granularities in Fig. 1 (a)-(c), and provide it.

3.2.2 Result
Fig. 2 shows the validity evaluation of four hour’s width set in the Figure 1(a)-(b). We obtain most evaluations to four hour’s width. Concerning the period, we hereafter adopt one week unit(Fig.1-(a)) in order to keep its recall, based on the discussion with the taxi company.
Fig. 2 Validity of Four Hour’s Width.

Fig. 3 Subjective Evaluation for Information Provision Sheet in the 1st Experiment.

Fig. 3 shows the result of the subjective evaluation for information provision sheet in the 1st experiment. Moreover, its sheet has brought us 4.1% time gain in waiting, and 16.7% one in cruising.

3.3 2nd Experiment: Investigation of Appropriate Listing Range of Picking-Up Numbers Ranking

3.3.1 Method

Table 2 Period of 2nd Experiment

<table>
<thead>
<tr>
<th>experiment period</th>
<th>one week from 2nd Sunday in August</th>
</tr>
</thead>
<tbody>
<tr>
<td>analyzed past data</td>
<td>from Aug. 10th (Sun.) to 16th (Sat.), 2008</td>
</tr>
<tr>
<td>recommendation execution</td>
<td>from Aug. 9th (Sun.) to 15th (Sat.), 2009</td>
</tr>
</tbody>
</table>

We change the listing range of picking-up numbers ranking in the information provision sheet from Top-3 to all that is equal to or larger than ten. After the 2nd recommendation experiment, we execute a questionnaire survey and ask to the taxi drivers: which range is useful for a taxi driver, 1st-10th, 11th-20th, 21st-30th, or from 31st to lower?

3.3.2 Result

Fig. 4 shows the evaluation result of useful range in ranking. Although the result shows that 1st-10th is most useful, it is possible to have significance in from 11th to lower. Therefore, we continue to list all that is equal to or larger than ten.

Fig. 4 Useful Ranking Range.

Fig. 5 Subjective Evaluation for Information Provision Sheet in the 2nd Experiment.

Fig. 5 shows the result of the subjective evaluation for the information provision sheet in the 2nd experiment. Moreover, its sheet has brought us 3.6% time gain in waiting, and 2.6% one in cruising.

3.4 3rd Experiment: Comparison of Importance between Picking-Up Numbers Information and Fare Achievement One

3.4.1 Method

Table 3 Period of 3rd Experiment

<table>
<thead>
<tr>
<th>experiment period</th>
<th>one week from 1st Sunday in Nov.</th>
</tr>
</thead>
<tbody>
<tr>
<td>analyzed past data</td>
<td>from Nov. 4th (Sun.) to 10th (Sat.), 2007, and from Nov. 2nd (Sun.) to 8th (Sat.), 2008</td>
</tr>
<tr>
<td>recommendation execution</td>
<td>from Nov. 1st (Sun.) to 7th (Sat.), 2009</td>
</tr>
</tbody>
</table>
In the 3rd experiment, adding to the picking-up numbers ranking per a location, we insert each mean fare achievement. After the 3rd recommendation experiment, we execute questionnaire survey and evaluate which has been more important, the numbers of picking-up or fare achievement?

3.4.2 Result
According to the Fig. 6, the numbers of picking-up is evaluated more important than fare achievement.

![Comparison of Importance](image)

Fig. 6 Which has been more important, the numbers of picking-up or fare achievement?

![Subjective Evaluation](image)

Fig. 7 Subjective Evaluation for Information Provision Sheet in the 3rd Experiment.

Fig. 7 shows the result of the subjective evaluation for the information provision sheet in the 3rd experiment. Moreover, its sheet has brought us 3.4% time gain in waiting, and 6.7% one in cruising.

3.5 4th Experiment: Investigation of Appropriate Granularity on Picking-Up Location

3.5.1 Method
We have used ‘address or institution name’ as the granularity of picking-up location from 1st to 3rd experiments. In the 4th experiment, we use broader granularity: ‘area’. After the 4th recommendation experiment, we execute questionnaire survey and evaluate which has been more useful, ‘address or institution name’ or ‘area’?

![Appropriate Granularity of Picking-Up Location](image)

Fig. 8 Appropriate Granularity of Picking-Up Location.

According to the Fig. 8, ‘area’ unit is evaluated more useful than ‘address or institution name’. Fig. 9 shows the result of the subjective evaluation for the information provision sheet in the 4th experiment. Moreover, its sheet has brought us 22.8% time gain in waiting, and 12.6% one in cruising.

![Subjective Evaluation](image)

Fig. 9 Subjective Evaluation for Information Provision Sheet in the 4th Experiment.

4 Concluding Remarks
In the present paper, we have analyzed occupied taxi data and tried to recommend promising ‘waiting/cruising’ location to taxi drivers. According
to the result of evaluation experiment, our proposition is effective.
As future research directions, we can point out the following two directions: (i) improvement of recommendation precision based on dynamic demands forecasting, (ii) effective collaboration with AVM system and/or taxi operator.

References: