Waiting/cruising location recommendation for efficient taxi business

T. Takayama, K. Matsumoto, A. Kumagai,
N. Sato, and Y. Murata

***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 our evaluation experiment, our proposition is effective.

***Keywords*—**Database, data mining, ITS(Intelligent Transport Systems), recommendation, and taxi.

I. Introduction

RECENTRY, 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 for 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. It visualizes and enables to manage location and status of each taxi in active, by the use of *GPS(Global Positioning System)*. Among the above-mentioned three types, it mainly supports 'wireless/order' method. However, we can not always say that there exists a sufficient method in order to support 'waiting' and/or 'cruising' types. 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.

The rest of the present paper is organized as follows. In section II, we summarize some previous works for efficient taxi business. Section III proposes some waiting/cruising location recommendation methods and investigates them through four experiments. Finally, in section IV, we conclude our paper and describe some future research directions.

II. Previous Works

In the paper[5], *Liao* surveys the method of AVM system collaborating with GPS receiver. It enables a taxi operator to efficiently allocate an empty taxi for an order from a passenger. It is used in 'wireless/order' method. *Hariharan et al.* try to represent and store trajectory from continuous tracking of location, not restricting to 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 a 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* tries to connect a taxi's supply chain as efficiently as possible[8]. If a taxi brings a passenger from location A to B, and after that, the same taxi brings another passenger from location B to C, then this taxi is not empty from location A to C. In the paper [9], *Lee et al.* analyze passenger picking-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 apply *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 wait for or cruise.

III. Waiting/Cruising Location Recommendation Based on Mining of Occupied Taxi Data

In this section, we describe our four experiments and try some recommendation methods.

A. Common Matters through Four Experiments

A.1 Analysis method

We analyze the occupied taxi data which each taxi driver records in the form of paper media, and find trends from them. Table 1 shows an example of occupied taxi data. One row corresponds to one trip. Concretely, 1st row shows that the driver #601 has brought a passenger from 'Yamagishi' to 'Chuo St.' on July 18th, 2008. Its occupied time is 8:25 AM. From 3rd experiments, we add 'fee achievement' column to the occupied taxi data. Wireless # shows the identifier of taxi.

Table 1 An Example of Occupied Taxi Data

data #	driver #	date	occupied time	source	destination	wireless #
1	601	July 18th, 2008	8:25	Yamagishi	Chuo St.	141
2	601	July 18th, 2008	8:50	Chuo St.	South-Senboku	141
3	601	July 18th, 2008	9:15	Nasukawa -town	Motomiya	141
…	…	…	…	…	…	…

Since there are some cases which a single taxi is shared by multiple drivers, each trip is recorded with both driver # and wireless #.

A.2 Recommendation method

We provide each taxi driver with location information where we recommend as waiting/cruising location using paper media which we call '*Information Provision Sheet; IPS*'. Concretely, we provide them with picking-up numbers ranking per location.

A.3 Evaluation method

We carry out the following three questionnaire surveys after each experiment:

(i) five levels subjective evaluation that the corresponding IPS has been useful or not. This five levels include 'useful', 'a little useful', 'neutral', 'little useful', and 'not useful'.

(ii) either they have changed their waiting/cruising location or not, based on the corresponding IPS.

(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.

A.4 Collaboration taxi company

We ask to provide us with occupied taxi data and to collaborate with our experiment to the company '*H*' in Morioka-City, which is the capital city of Iwate prefecture in Japan. The population of this city is about 300 thousands. The company 'H' has 100 taxis and 150 taxi drivers.

B 1st Experiment: Investigation of Time Granularity

B.1 Method

Before the execution of recommendation experiment, we conduct some opinion polls concerning waiting/cruising. Their questions are shown in Fig. 1. Questions about waiting are Q1 and Q2, and ones about cruising are Q3 and Q4.

After this opinion poll, we carry out our 1st recommendation experiment shown in Table 2. In determination of its experiment period, we avoid a big social event and reduce an influence by a particular event.

Q1. Do you have confidence concerning waiting location?
- Yes
- a little Yes
- neutral
- a little No
- No
- others

Q2. Do you need more information concerning waiting location?
- Yes
- a little Yes
- neutral
- a little No
- No
- others

Q3. Do you have confidence concerning cruisting location?
- Yes
- a little Yes
- neutral
- a little No
- No
- others

Q4. Do you need more information concerning cruising location?
- Yes
- a little Yes
- neutral
- a little No
- No
- others

Fig. 1 Opinion Poll concerning Waiting/Cruising.

Table 2 Period of 1st Experiment

experiment period	one week from 3rd Fri. in Jul.
analyzed past data	from Jul. 18th (Fri.) to 24th(Thu.), 2008
recommendation execution	from Jul. 17th (Fri.) to 23th(Thu.), 2009

Based on the discussion with the taxi company, we prepare three types of time granularity as shown in Fig. 2. In Fig. 2-(a), we make six blocks. Each block includes four hours through one week. This means that one block has twenty-eight hours because four hours multiplied by seven days are equal to twenty-eight hours. In Fig.2-(b), each block has just four hours because each block includes only one day. In Fig.2-(c), one block corresponds to working shift through one week. Actually, there are seven types of working shift. A driver works from morning to evening, another driver works from night to morning, and so on. However, in order to save a space, we write only one shift of them in this figure. Towards each block, we make picking-up numbers Top-3 ranking in address unit.

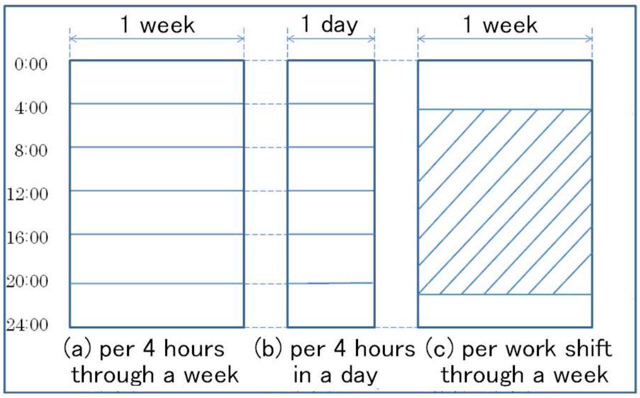

Fig.2 Three Alternatives of Time Granularity.

Table 3 shows an example of IPS. It is based on Fig. 2-(a). In seven days of our 1st experiment, we use two days for each try. On only the middle day, we provide with no IPS.

We carry out questionnaire survey after the week of our 1st experiment is over. Its questions are shown in Fig. 3.

Q1. Did you change your waiting/cruising location?
- waiting changed
- cruising changed
- both changed
- both not changed

Q2. Which time granularity is useful?
- per four hours through a week
- per four hours in a day
- per work shift through a week
- nothing useful

Q3. Is four hours granularity valid?
- too wide
- too narrow
- valid
- others(concretely)

Q4. When you had tried 'waiting' in the 1st experiment period, how long was the Mean Time until Occupied? (Except for the case which you had received a wireless order)
- equal to or less than twenty minutes
- from twenty to forty minutes
- from forty to sixty minutes
- from sixty to eighty minutes
- equal to or larger than eighty minutes

Q5. When you had tried 'cruising' in the 1st experiment period, how long was the Mean Time until Occupied? (Except for the case which you had received a wireless order)
- equal to or less than twenty minutes
- from twenty to forty minutes
- from forty to sixty minutes
- from sixty to eighty minutes
- equal to or larger than eighty minutes

Fig. 3 Questionnaire Survey after Our 1st Experiment Week is Over.

Table 3 An Example of IPS per Four Hours through One Week Granularity.

ranking / time period	rank 1		rank 2		rank 3	
	location	numbers of picking up	location	numbers of picking up	location	numbers of picking up
0:00 to 4:00	Oodori	127	Honsya	50	Uchimaru	30
4:00 to 8:00	Morioka Station	65	Yamagishi	56	Tsushida	32
8:00 to 12:00	Yamagishi	272	Morioka Station	200	Tsushida	76
12:00 to 16:00	Morioka Station	223	Yamagishi	201	Chuo St.	56
16:00 to 20:00	Morioka Station	258	Yamagishi	169	Oodori	100
20:00 to 24:00	Oodori	176	Morioka Station	167	Honsya	94

B.2 Result

We start from the result of the opinion poll. The numbers of response is 66, and its collection rate is 44.0%. Table 4 shows the confidence concerning their waiting location. Positive answers including ‘Yes’ and ‘a little Yes’ are 37.9%. On the other hand, negative answers including ‘No’ and ‘a little No’ are 16.7%.

Table 4 Confidence concerning Waiting Location.

answer	count	%
Yes	10	15.2
a little Yes	15	22.7
neutral	29	43.9
a little No	7	10.6
No	4	6.1
others	1	1.5
sum	66	100

Table 5 shows the result of “Do you need more information concerning waiting location?” Positive answers including ‘Yes’ and ‘a little Yes’ are 59.1%. On the other hand, negative answers including ‘No’ and ‘a little No’ are 13.6%.

Table 5 Do You Need More Information concerning Waiting Location?

answer	count	%
Yes	17	25.8
a little Yes	22	33.3
neutral	16	24.2
a little No	8	12.1
No	1	1.5
others	2	3.0
sum	66	100

From the result of Table 4 and 5, we can say that the needs for waiting location information is not small.

Table 6 shows the confidence concerning their cruising location, and Table 7 shows the result of “Do you need more information concerning cruising location?” From the result of Table 6 and 7, we can say that the needs for cruising location information is not small, neither.

Table 6 Confidence of Cruising Location.

answer	count	%
Yes	6	9.1
A little Yes	12	18.2
neutral	32	48.5
A little No	13	19.7
No	0	0
others	3	4.5
sum	66	100

Table 7 Do You Need More Information concerning Cruising Location?

answer	count	%
Yes	17	25.8
a little Yes	18	27.3
neutral	20	30.4
a little No	6	9.1
No	3	4.5
others	2	3.0
sum	66	100

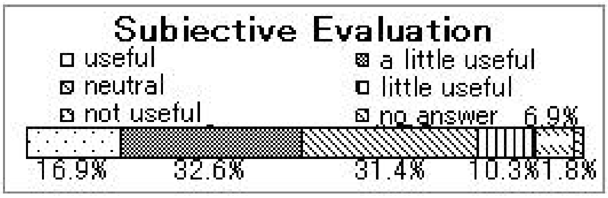

Fig. 4 Subjective Evaluation for IPS in the 1st Experiment.

Fig. 4 shows the result of the subjective evaluation for IPS in the 1st experiment. The sum of the positive evaluation including ‘useful’ and ‘a little useful’ is 49.5%. On the other hand, the sum of the negative evaluation including ‘little useful’ and ‘not useful’ is 17.2%. It means that positive evaluation is larger than twice of negative one.

Table 8 shows the answer for the question: “Which time granularity is appropriate?” Although ‘work shift through 1 week’ obtains the best evaluation, its boundaries change depending upon the company. It means the lack of generality. Furthermore, a single work shift often includes morning and night. In general, trend in morning is different from one in night. We can not always say that we adopt a block including morning and night as a time granularity

The second best is ‘per 4 hours in 1 day’. However, we have only one year data. If we adopt this granularity, we have to forecast a certain day trend from only one day, the same day of the one year before. Therefore, we can not always say that we have enough data to forecast. In order to adopt ‘per 4 hours in 1 day’, we should have at least five years data. Based on the discussion with the taxi company, concerning the period, we hereafter adopt one week unit(Fig.2-(a)) in order to keep its recall property. This judgment is temporal, and not perfect. If we could obtain enough data for forecasting, ‘per 4 hours in 1 day’ is preferable. However, since the volume of occupied taxi data is in general huge, we can not always say that many taxi companies save their occupied data in many years. Even in the case that the AVM system is used, in general, their data are sensor data, huge and put away without saving them in the computer. In summary, as a realistic judgment, we adopt ‘4 hours through 1 week’ as time granularity.

Fig. 5 shows the validity evaluation of four hour’s width set in the Figure 2(a)-(b). We obtain the best evaluations to four hour’s width.

Table 8 Which time granularity is appropriate?

answer	count	%
per 4 hours through 1 week	13	19.7
per 4 hours in 1 day	16	24.2
Work shift through 1 week	25	37.9
nothing useful	10	15.2
no answer	2	3.0
Sum	66	100.0

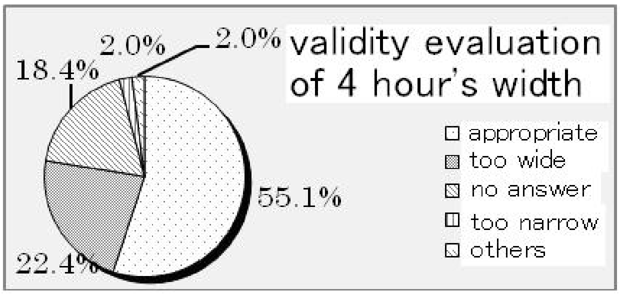

Fig. 5 Validity of Four Hour's Width.

Fig. 6 shows the achievement of location movement in the 1st experiment. The numbers of response is 49, and its collection rate is 32.7%. Sum of the response 'waiting changed', 'cruising changed', and 'both changed' is 38.8%. It is smaller than 53.1% of the ratio: 'not changed'. We consider its reason as follows:

- Since we make the picking up numbers ranking at top three, their locations in the IPS are not much different from their ordinal waiting or cruising locations.
- Another possibility is that the taxi drivers do not like to change their favorite waiting/cruising location.

We have obtained a common opinion from some drivers. "Ranking range of top three is too narrow". From the 2nd experiment, we enlarge its ranking range and try to resolve this 'too narrow' problem.

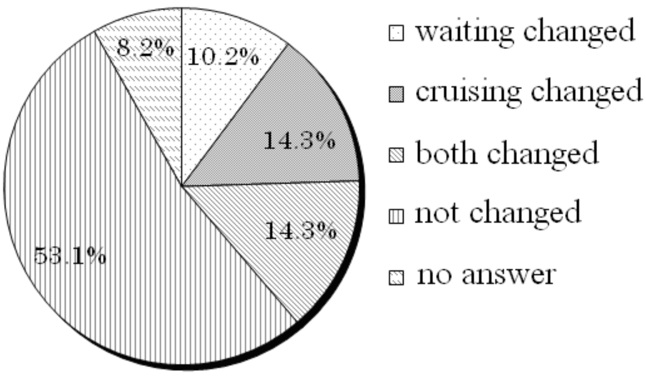

Fig. 6 Achievement of Location Movement in the 1st Experiment.

Table 9 shows *Mean Time until Occupied(MTO)* in the 1st experiment. From the answer of Q4 and Q5 in the Fig. 3, we calculate MTO. We adopt each representative value for each answer. Concretely, we treat 'equal to or less than twenty minutes' as 'ten minutes', 'from twenty to forty minutes' as 'thirty minutes', 'from forty to sixty minutes' as 'fifty minutes', and son on. Concerning 'equal to or larger than eighty minutes', we treat it as 'ninety minutes'. We have obtained gain: 2.5 minutes in waiting and 10.0 minutes in cruising. In other words, the IPS in the 1st experiment has brought us 4.1% time gain in waiting, and 16.7% one in cruising.

Table 9 MTO in the 1st Experiment

driver's behavior	MTO(min.)
(a) all drivers who change the waiting location	58.3
(b) all drivers who change the cruising location	50.0
(c) all drivers who do not change the waiting location	60.8
(d) all drivers who do not change the cruising location	60.0
(e) waiting time reduction {(c)-(a)}	2.5
(f) cruising time reduction {(d)-(b)}	10.0

Concerning the granularity of four hours through one week, we have investigated the recall property. Concretely, we have used the picking up numbers ranking data in 2008 and 2009. Based on the *Spearman's ranking correlation*[11], we calculate correlation coefficient between them, and write its scatter diagram.

Table 10 shows its result. Correlation coefficient is from 0.66 to 0.87. It means a little or strong correlation. Fig. 7shows a scatter diagram in the case from 8:00 to 12:00.

Table 10 Correlation Coefficient between 2008 and 2009

time period	correlation coefficient
0:00 to 4:00	0.75
4:00 to 8:00	0.66
8:00 to 12:00	0.87
12:00 to 16:00	0.76
16:00 to 20:00	0.87
20:00 to 24:00	0.81

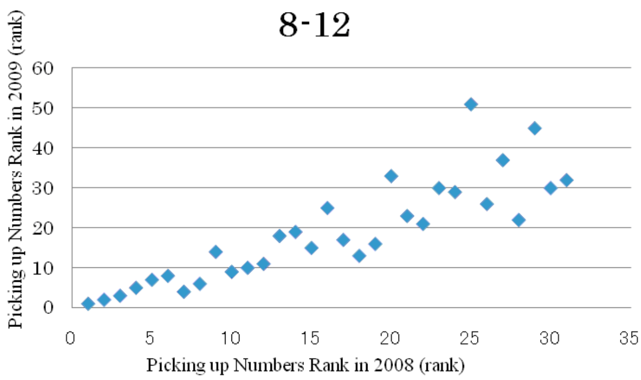

Fig. 7 Scatter Diagram of Ranking for Picking up Numbers from 8:00 to 12:00 between 2008 and 2009

It is right-upper. Other time range has almost the same trend. We can consider that the picking up numbers ranking has recall property beyond year. Therefore, we can say that it is valid to recommend the waiting/cruising location based on the past picking up record.

C 2nd Experiment: Investigation of Appropriate Listing Range of Picking up Numbers Ranking

C.1 Method

We carry out our 2nd recommendation experiment shown in Table 11. In the 1st experiment, we can not say that the interest to our experiment in subject taxi drivers is sufficient. Based on the discussion with the taxi company, we set the experiment period in our 2nd experiment to ‘one week from 2nd Sunday in August’. Most of this week is summer vocation in our country. Since the taxi drivers are not busy in this period, they can carefully watch our IPS.

Table 11 Period of 2nd Experiment

experiment period	one week from 2nd Sun. in Aug.
analyzed past data	from Aug. 10th (Sun.) to 16th (Sat.), 2008
recommendation execution	from Aug. 9th (Sun.) to 15th (Sat.), 2009

We try to solve the problem in the 1st experiment that the listing range of picking up numbers ranking in the IPS is too narrow. Concretely, we change the listing range from Top-3 to all that is equal to or larger than ten. Moreover, in the 1st experiment, we have provided the taxi drivers with one of the three types of IPSs depending upon each day. From the 2nd experiment, we provided them just the one type IPS, ‘per four hours through one week’, on the start day of the experiment. The sample of this IPS is like Fig. 12. In each four hours, we write ‘address or institution name’ and ‘picking up numbers’ which is larger than ten.

Table 12 A Sample of IPS in the 2nd Experiment

0:00 to 4:00 picking up numbers ranking

rank	picking up		rank	picking up	
	location	#		location	#
1	Oodori	69	7	Sanbon-Yanagi	16
2	Iwate Nippou company	31	8	Honcho St.	14
3	Uchimaru	28	9	Yamagishi	13
4	Honsya	26	10	Kita-Ohashi	13
5	Tsushida	21	11	Saien	12
6	Chuo St.	21	12	Takamatsu	11

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? Fig. 8 shows its questionnaire survey in our 2nd experiment. The differences from one in the 1st experiment are following two points:

1) we change Q2 into a question where we ask the subjective usefulness of the corresponding IPS.
2) we change Q3 into the question that which range is most useful for you.

Other questions are the same as ones in the case of 1st experiment.

Q1. Did you change your waiting/cruising location?
- waiting changed
- cruising changed
- both changed
- both not changed

Q2. How about the subjective usefulness of the corresponding IPS?
- useful
- a little useful
- neutral
- little useful
- not useful

Q3. Which range is most useful for you?
- 1st to 10th
- 11th to 20th
- 21st to 30th
- 31st or more larger
- not useful at all

Q4. When you had tried ‘waiting’ in the 2nd experiment period, how long was the Mean Time until Occupied? (Except for the case which you had received a wireless order)
- equal to or less than twenty minutes
- from twenty to forty minutes
- from forty to sixty minutes
- from sixty to eighty minutes
- equal to or larger than eighty minutes

Q5. When you had tried ‘cruising’ in the 2nd experiment period, how long was the Mean Time until Occupied? (Except for the case which you had received a wireless order)
- equal to or less than twenty minutes
- from twenty minutes to forty minutes
- from forty minutes to sixty minutes
- from sixty minutes to eighty minutes
- equal to or larger than eighty minutes

Fig. 8 Questionnaire Survey in Our 2nd Experiment.

C.2 Result

Fig. 9 shows the evaluation result of useful range in ranking. The result shows that 1st-10th is most useful. Although the listing range in the 1st experiment is too narrow, too many listing is also not adequate. It has possibility to have significance from 11th to lower. Based on the discussion with the taxi company, we hereafter continue to list all that is equal to or larger than ten.

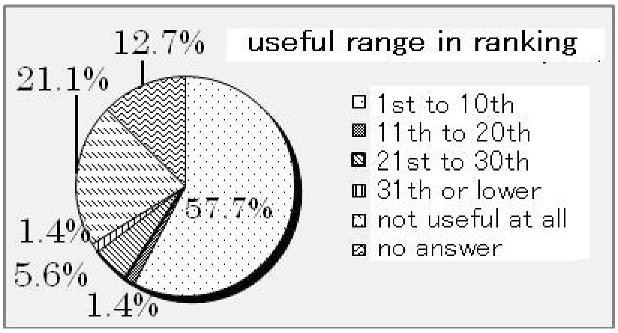

Fig. 9 Useful Ranking Range.

Fig. 10 shows the result of the subjective evaluation for the IPS in the 2nd experiment. We can observe that the answer 'a little useful' is increasing more than the 1st experiment. On the other hand, we can observe that the answer 'useful' is decreasing. As we have described in the Table 11, most of the period in the 2nd experiment is summer vocation. Its picking up numbers is decreasing itself, and it has possibility to affect the subjective evaluation.

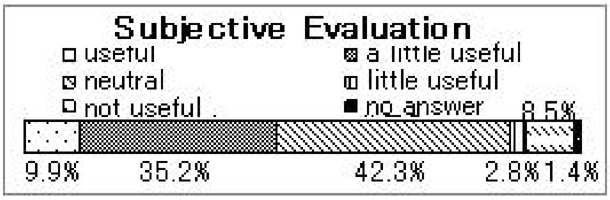

Fig. 10 Subjective Evaluation for IPS in the 2nd Experiment.

Fig. 11 shows the achievement of location movement in the 2nd experiment. As a whole, we can say that the change from the 1st experiment is not large.

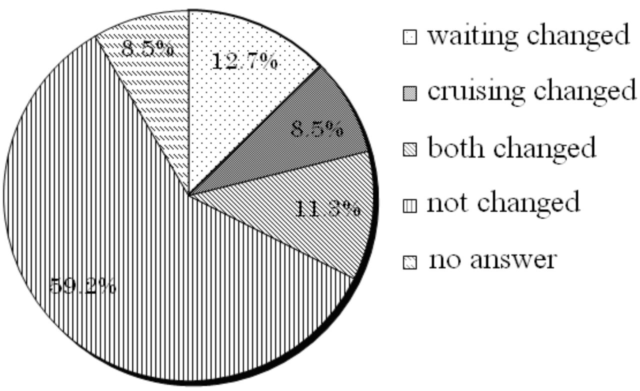

Fig. 11 Achievement of Location Movement in the 2nd Experiment.

Table 13 shows MTO in the 2nd experiment. We have obtained a little gain: 2.1 minutes in waiting and 1.5 minutes in cruising. In other words, our IPS has brought us 3.6% time gain in waiting, and 2.6% one in cruising.

Table 13 MTO in the 2nd Experiment

driver's behavior	MTO(min.)
(a) all drivers who change the waiting location	51.2
(b) all drivers who change the cruising location	55.0
(c) all drivers who do not change the waiting location	53.1
(d) all drivers who do not change the cruising location	56.5
(e) waiting time reduction {(c)-(a)}	2.1
(f) cruising time reduction {(d)-(b)}	1.5

D 3rd Experiment: Comparison of Importance between Picking up Numbers and Fee Achievement

D.1 Method

We carry out our 3rd recommendation experiment shown in Table 14. This period does not overlap some vocation like the 2nd experiment. Furthermore, as in the case of the 1st experiment, its period does not overlap some big social events, and an influence by a particular event is not much.

Table 14 Period of 3rd Experiment

experiment period	one week from 1st Sun. in Nov.
analyzed past data	from Nov. 4th(Sun.) to 10th(Sat.), 2007, and from Nov. 2nd(Sun.) to 8th(Sat.), 2008
recommendation execution	from Nov. 1st(Sun.) to 7th(Sat.), 2009

In the 3rd experiment, adding to the picking up numbers ranking per a location, we insert each fee achievement. Concretely, we add mean, maximum, and minimum fee achievement per a location. From this 3rd experiment, we have been able to obtain picking up numbers data not only in 2008 year but also in 2007 one. We use these two year data and as in the case of the 2nd experiment, we make our IPS of picking up numbers ranking per four hours through one week, under the condition that its picking up numbers is equal or more than ten. First, we make its ranking based on the data in 2008. Second, concerning the location in the 2008 ranking, we add the data in 2007. Fig. 12 shows a sample of this IPS. Adopting a proposition from the taxi company, from this IPS, we also write picking up numbers per one hour.

After the 3rd recommendation experiment, we execute questionnaire survey and evaluate which has been more important, the numbers of picking up or fee achievement? Fig. 13 shows its questionnaire survey in our 3rd experiment. The differences from one in the 2nd experiment are as follows:

- we change Q3 into the question that asks which is more important, the picking up numbers or fee achievement?

Other questions are the same as ones in the case of 2nd experiment.

Picking up Numbers Ranking from 0:00 to 4:00

one week from 1st Sun. in Nov., 2008					picking up location	one week from 1st Sun. in Nov., 2007				
cf. fee achievement			picking up numbers (per hour)	rank		rank	picking up numbers (per hour)	cf. fee achievement		
max.	mean	min.						min.	mean	max.
5880	1319	520	172(6.14)	1	Oodori	1	114(4.07)	520	1401	7240
5160	1377	520	39(1.39)	2	Chuo St.	2	40(1.43)	520	1360	6760
2360	1232	520	33(1.18)	3	Honsya	3	21(0.75)	520	1419	4920
1880	954	520	21(0.75)	4	Saien	5	20(0.71)	520	1215	2280
2440	1096	520	20(0.71)	5	Uchimaru	5	20(0.71)	520	1924	4840
2200	844	520	19(0.68)	6	Honcho St.	7	17(0.61)	520	891	2280
3640	2094	600	19(0.68)	6	Iwate Nippou Inc.	3	21(0.75)	600	1796	3480
2600	1525	520	14(0.50)	8	Sanbonyanagi	11	8(0.29)	680	1750	3240
4520	1492	600	13(0.46)	9	Tsushida	8	14(0.50)	600	1312	2440
2440	1168	520	10(0.36)	10	Sakanacho	18	5(0.18)	600	1608	3000
1160	1000	760	10(0.36)	10	Yamagishi	10	10(0.36)	600	1112	2600

Fig. 12 A Sample of IPS in the 3rd Experiment.

Q1. Did you change your waiting/cruising location?
- waiting changed
- cruising changed
- both changed
- both not changed

Q2. How about the subjective usefulness of the corresponding IPS?
- useful
- a little useful
- neutral
- little useful
- not useful

Q3. Which is more important for you, picking up numbers or fee achievement ?
- picking up numbers
- fee achievement
- neutral
- unknown

Q4. When you had tried 'waiting' in the 3rd experiment period, how long was the Mean Time until Occupied? (Except for the case which you had received a wireless order)
- equal to or less than twenty minutes
- from twenty to forty minutes
- from forty to sixty minutes
- from sixty to eighty minutes
- equal to or larger than eighty minutes

Q5. When you had tried 'cruising' in the 3rd experiment period, how long was the Mean Time until Occupied? (Except for the case which you had received a wireless order)
- equal to or less than twenty minutes
- from twenty to forty minutes
- from forty to sixty minutes
- from sixty to eighty minutes
- equal to or larger than eighty minutes

Fig. 13 Questionnaire Survey in Our 3rd Experiment.

We also investigate the recall property using two years data. Its method is basically the same as one in the 1st experiment. First, we adopt the data in 2008 and 2009. We calculate correlation coefficient between them, and write its scatter diagram. Second, we adopt the data in 2007 and 2009. On the same way, we calculate correlation coefficient between them, and write its scatter diagram.

D.2 Result

According to the Fig. 14, the picking up numbers is evaluated more important than fee achievement.

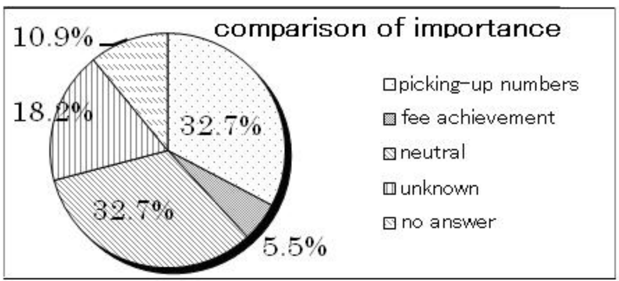

Fig. 14 Which Has Been More Important, the Picking up Numbers or Fee Achievement?

We continue to use fee achievement information based on the following two reasons:

- The questionnnaire survey in the 3rd experiment is relative comparison between 'picking up numbers' and 'fee achievement'. Even if a subject answers "picking up numbers are more important", it does not mean that "fee achievement is not useful at all".
- Actually, we have obtained the answer that "fee achievement is more important than picking up numbers" at 5.5 % ratio.

Fig. 15 shows the result of the subjective evaluation for the IPS in the 3rd experiment. We can classify the answer 'useful' and 'a little useful' into positive ones, and 'little useful' and 'not useful' into negative ones. According to the Fig. 14, its change from the 2nd experiment is not large.

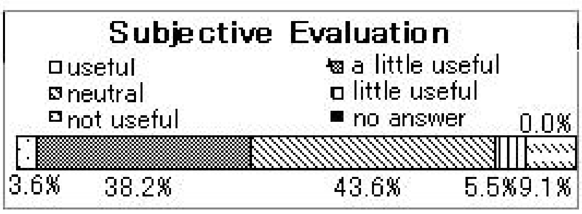

Fig. 15 Subjective Evaluation for IPS in the 3rd Experiment.

Fig. 16 shows the achievement of location movement in the 3rd experiment. Since we have obtained the corresponding answers from fifty-five taxi drivers, its collection rate is 36.7%. 'Cruising changed' increases 6.0 % from the 2nd experiment.

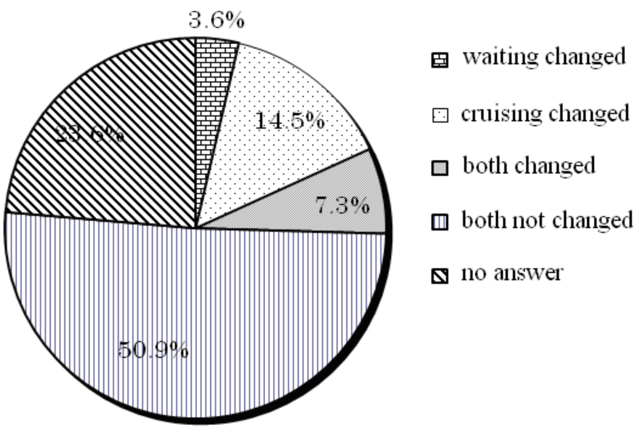

Fig. 16 Achievement of Location Movement in the 3rd Experiment.

Table. 15 shows MTO in the 3rd experiment. We have obtained a little gain: 1.8 minutes in waiting and 4.0 minutes in cruising. In other words, our IPS has brought us 3.4% time gain in waiting, and 6.7% one in cruising.

Table 16 shows the correlation coefficients between 2008 and 2009 picking up numbers ranking. Their values are from 0.71 to 0.90, and it means strong correlation. Table 17 shows the case between 2007 and 2009. We have obtained almost the same results. We can say that picking up numbers ranking is recalled beyond year.

Table 15 MTO in the 3rd Experiment

driver's behavior	MTO(min.)
(a) all drivers who change the waiting location	50.0
(b) all drivers who change the cruising location	55.5
(c) all drivers who do not change the waiting location	51.8
(d) all drivers who do not change the cruising location	59.5
(e) waiting time reduction {(c)-(a)}	1.8
(f) cruising time reduction {(d)-(b)}	4.0

Table 16 Correlation Coefficient between 2008 and 2009

time period	correlation coefficient
0:00 to 4:00	0.82
4:00 to 8:00	0.71
8:00 to 12:00	0.78
12:00 to 16:00	0.80
16:00 to 20:00	0.75
20:00 to 24:00	0.90

Table 17 Correlation Coefficient between 2007 and 2009

time period	correlation coefficient
0:00 to 4:00	0.74
4:00 to 8:00	0.82
8:00 to 12:00	0.86
12:00 to 16:00	0.78
16:00 to 20:00	0.83
20:00 to 24:00	0.77

Fig. 17 shows a scatter diagram in the case from 8:00 to 12:00 in 2008 and 2009. It is right-upper. Other time range has almost the same trend. Fig. 18 shows a scatter diagram in the case from 8:00 to 12:00 in 2007 and 2009. It is also right-upper. Other time range has almost the same trend. We can consider that the picking up numbers ranking has also recall property beyond year.

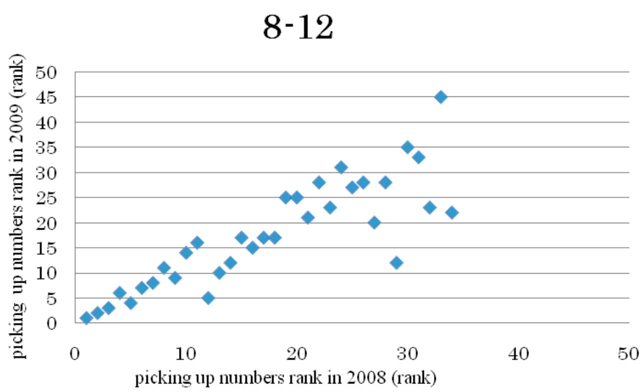

Fig. 17 A Sample of Scatter Diagram between 2008 and 2009.

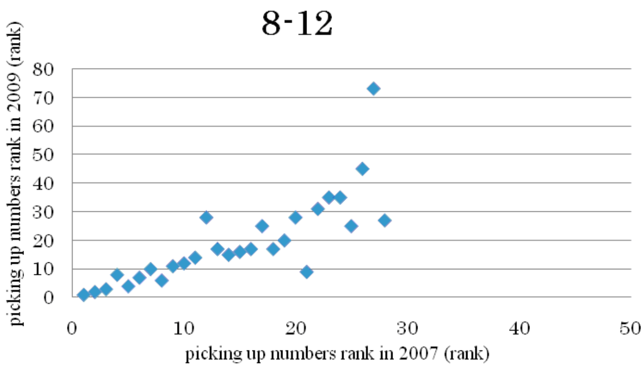

Fig. 18 A Sample of Scatter Diagram between 2007 and 2009.

E 4th Experiment: Investigation of Appropriate Granularity on Picking up Location

E.1 Method

We carry out our 4th recommendation experiment shown in Table. 18. This period does not overlap some vocation like the 2nd experiment, neither. Furthermore, as in the case of the 1st and 3rd experiment, its period does not overlap some big social events, and an influence by a particular event is not much.

We have used 'address or institution name' as the granularity of picking up location from 1st to 3rd experiments. In the 4th experiment, we introduce a broader granularity: 'area'. We classify each address or institution name into one of the eight areas. We make picking up numbers ranking at area unit, and use it as IPS.

Table 18 Period of 4th Experiment

experiment period	one week from 4th Sun. in Nov.
analyzed past data	from Nov. 25th(Sun.) to Dec. 1st(Sat.), 2007 from Nov. 23rd(Sun.) to 29th(Sat.), 2008
recommendation execution	from Nov. 22nd(Sun.) to 28th(Sat.), 2009

Fig. 19 shows a sample of this IPS. Except for location granularity, its content is the same as the case in the 3rd experiment.

After the 4th recommendation experiment, we execute questionnaire survey and evaluate which granularity has been more useful, 'address or institution name' or 'area'?

Fig. 20 shows its questionnaire survey in our 4th experiment. The differences from one in the 3rd experiment are as follows:

- we change Q3 into the question that asks which granularity is more appropriate as picking up location, 'address or institution name' or 'area'?

Other questions are the same as ones in the case of 3rd experiment.

We also investigate the recall property using two years data. Its method is basically the same as one in the 3rd experiment. First, we adopt the data in 2008 and 2009. We calculate correlation coefficient between them, and write its scatter diagram. Second, we adopt the data in 2007 and 2009. On the same way, we calculate correlation coefficient between them, and write its scatter diagram.

Picking up Numbers Ranking from 0:00 to 4:00

one week from 4th Sun. in Nov., 2008					picking up area	one week from 4th Sun. in Nov., 2007				
cf. fee achievement			picking up numbers (per hour)	rank		rank min.	picking up numbers (per hour) mean	cf. fee achievement		
max.	mean	min.						max.	mean	max.
5480	1261	520	451(16.11)	1	Honsya	1	280(10.00)	520	1298	5400
4600	1091	520	48(1.71)	2	Kita-Ohashi	3	41(2.04)	520	1308	4120
2440	1008	520	40(1.43)	3	Minami-Oodori	4	31(1.46)	520	1258	4680
3220	1277	520	38(1.36)	4	Tonan	2	57(1.11)	520	1678	6840
2680	1224	520	25(0.89)	5	Morioka Sta. front	5	17(0.61)	520	1531	4760
1640	1096	680	10(0.36)	6	Yamagishi	6	15(0.54)	520	1069	2200
1640	1266	1000	6(0.21)	7	Seinan	7	4(0.14)	1080	1660	3240
1720	1720	1720	1(0.04)	8	Others	8	3(0.11)	1080	3320	6040

Fig. 19 A Sample of IPS in the 4th Experiment.

Q1. Did you change your waiting/cruising location?
- waiting changed
- cruising changed
- both changed
- both not changed

Q2. How about the subjective usefulness of the corresponding IPS?
- useful
- a little useful
- neutral
- little useful
- not useful

Q3. Which granularity is more appropriate as picking up location, 'address or institution name' or 'area'?
- address or institution name
- area
- neutral
- unknown

Q4. When you had tried 'waiting' in the 4th experiment period, how long was the Mean Time until Occupied? (Except for the case which you had received a wireless order)
- equal to or less than twenty minutes
- from twenty to forty minutes
- from forty to sixty minutes
- from sixty to eighty minutes
- equal to or larger than eighty minutes

Q5. When you had tried 'cruising' in the 4th experiment period, how long was the Mean Time until Occupied? (Except for the case which you had received a wireless order)
- equal to or less than twenty minutes
- from twenty minutes to forty minutes
- from forty minutes to sixty minutes
- from sixty minutes to eighty minutes
- equal to or larger than eighty minutes

Fig.20 Questionnaire Survey in Our 4th Experiment.

E.2 Result

Since we have obtained the corresponding answers from forty-nine taxi drivers, its collection rate is 32.7%. According to the Fig. 21, 'area' unit is evaluated more appropriate than 'address or institution name'. Based on the discussion with the taxi company, its reason is as follows. When a taxi driver moves their waiting/cruising location based on the IPS, conventional IPS at address or institution name unit has too many candidate locations. It is not easy for them to determine which location they should go. On the other hand, area unit includes only eight candidates, and it is more easy for them to determine it. Moreover, its spatial size is more appropriate. Especially, in cruising, if a recommended location is too narrow, it easily happens to leave the recommended location by cruising.

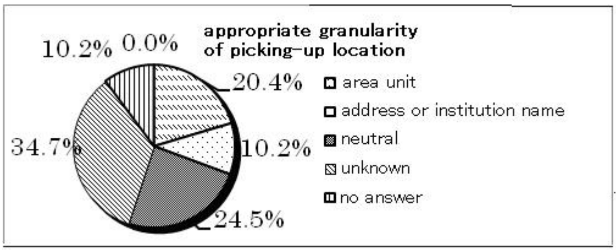

Fig. 21 Appropriate Granularity of Picking-Up Location.

Fig. 22 shows the result of the subjective evaluation for the IPS in the 4th experiment. The answer 'neutral' is 51.0 % and increases obviously. We are required to keep their interest in more fresh.

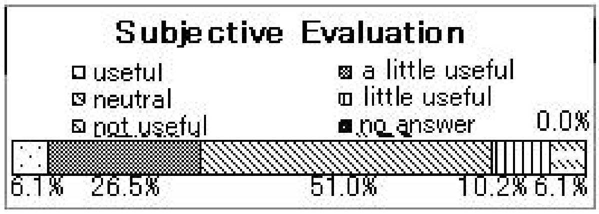

Fig. 22 Subjective Evaluation for IPS in the 4th Experiment.

Fig. 23 shows the achievement of location movement in the 4th experiment. 'Waiting changed' increases 10.7 % from the 3rd experiment. 'Not changed' decreases 8.0% and it means that IPS in the 4th experiment is used in active.

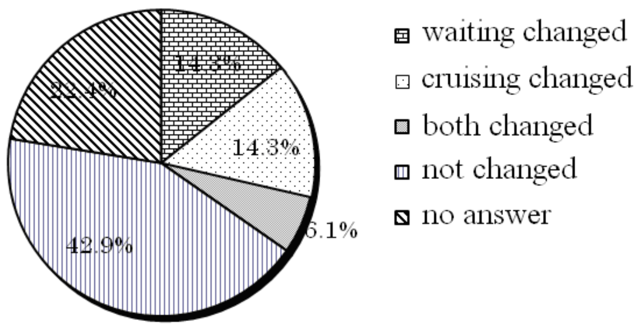

Fig. 23 Achievement of Location Movement in the 4th Experiment.

Table 19 shows MTO in the 4th experiment. We have obtained gain: 13.0 minutes in waiting and 7.5 minutes in cruising. In other words, the IPS has brought us 22.8% time gain in waiting, and 12.6% one in cruising.

Table 20 shows the correlation coefficients between 2008 and 2009 picking up numbers ranking. In the range from 8:00 to 24:00, their values are from 0.80 to 0.90, and it means strong correlation. Table 21 shows the case between 2007 and 2009. We have obtained almost the same results. In this period, we can say that picking up numbers ranking is recalled beyond year. On the other hand, in the period from 0:00 to 8:00, we have obtained relatively less values: from 0.50 to 0.67 in two years. These values mean a little correlation.

Table 19 MTO in the 4th Experiment

driver's behavior	MTO(min.)
(a) all drivers who change the waiting location	44.0
(b) all drivers who change the cruising location	52.0
(c) all drivers who do not change the waiting location	57.0
(d) all drivers who do not change the cruising location	59.5
(e) waiting time reduction {(c)-(a)}	13.0
(f) cruising time reduction {(d)-(b)}	7.5

Table 20 Correlation Coefficient between 2008 and 2009

time period	correlation coefficient
0:00 to 4:00	0.64
4:00 to 8:00	0.64
8:00 to 12:00	0.80
12:00 to 16:00	0.90
16:00 to 20:00	0.89
20:00 to 24:00	0.82

Table 21 Correlation Coefficient between 2007 and 2009

time period	correlation coefficient
0:00 to 4:00	0.50
4:00 to 8:00	0.67
8:00 to 12:00	0.77
12:00 to 16:00	0.79
16:00 to 20:00	0.85
20:00 to 24:00	0.91

Fig. 24 shows a scatter diagram in the case from 8:00 to 12:00 in 2008 and 2009. It is right-upper. Other time range has almost the same trend. Fig. 25 shows a scatter diagram in the case from 8:00 to 12:00 in 2007 and 2009. It is also right-upper. Other time range has almost the same trend.

We can consider that the picking up numbers ranking in the case of area unit has also recall property beyond year. Therefore, we can say that it is valid to recommend the waiting/cruising location based on the past picking up record at area unit.

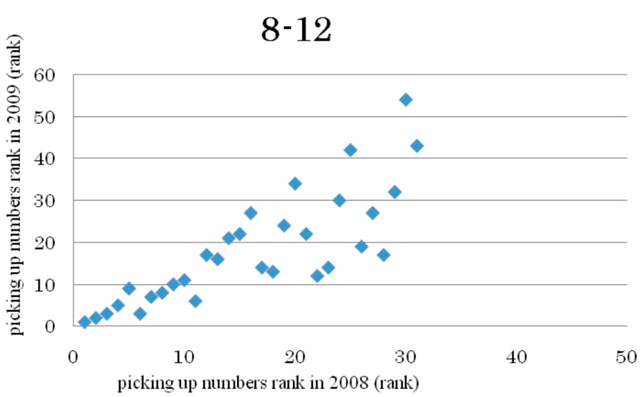

Fig. 24 Scatter Diagram of Ranking for Picking up Numbers from 8:00 to 12:00 between 2008 and 2009.

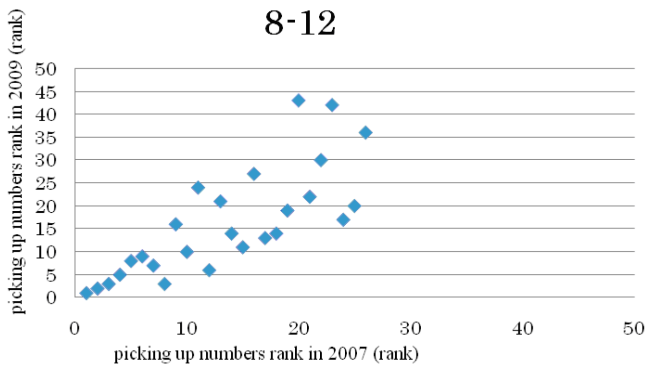

Fig. 25 Scatter Diagram of Ranking for Picking up Numbers from 8:00 to 12:00 between 2007 and 2009.

IV. 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 our evaluation experiment, our proposition is effective.

The present paper has collaborated with a taxi company *'H'* in our city. We have used about 70 thousands of occupied taxi data from 2007 to 2009 in this paper. Its volume is not small. However, it is preferable to collaborate with multiple taxi companies. It is the common problem to the paper [9]. Since the other taxi company except for 'H' is competitor in the same city, it is not realistic to collaborate with them. However, if we collaborate with a taxi company in a far city, it comes to difficult to execute our experiments. Now we are looking for a good collaborator with us.

As future research directions, we can point out the following three directions: (i)improvement of recommendation precision based on dynamic demands forecasting, (ii)effective collaboration with AVM system and/or taxi operator, and (iii)evaluation with other taxi company.

Acknowledgment

We would like to thank the taxi company *'H'* in Morioka-City, Japan.

References

[1] Z. W. Li, M. Zhu: "Breadth-first search based bus transport transfer algorithm," in *Proc. of the 9th WSEAS International Conference on Applied Computer and Applied Computational Science*, 2010, pp.106-110.
[2] P. Bojar, Ł. Muślewski, M. Woropay, J. Szpytko: "Influence of human factor on transport system safety," in *Proc. of the WSEAS 4th International Conference on Applied Mathematics, SimulationI, Modelling*, 2010, pp.227-231.
[3] D. Peidro, M. D. Madronero, J. Mula: "Operational transport planning in an automobile supply chain: an interactive fuzzy multi-objective approach," in *Proc. of the 8th WSEAS International Conference on Computational Intelligence, Man-Machine Systems and Cybernetics*, 2009, pp.121-127.
[4] T. Zelinka, M. Svitek: "Adaptive communications solutions in complex transport telematics systems," in *Proc. of the 12th WSEAS International Conference on Communications*, 2008, pp.206-212.
[5] Z. Liao: "Real-time taxi dispatching using global positioning systems," *Communications of the ACM*, vol.46, no.5, 2003, pp.81-83.

[6] R. Hariharan, K. Toyama: “Project lachesis: parsing and modeling location histories,” in *Proc. of the GIScience2004*, LNCS, vol.3234, 2004, pp.106-124.
[7] A. P. Silva: “Location-based taxi service in wireless communication environment,” in *Proc. of the IEEE 36th Annual Simulation Symposium(ANSS’03)*, 2003, pp.47-54, 2003.
[8] S. T. Hou: “Comfort taxi: managing service supply chain in transportation industry,” in *Proc. of the IEEE International Symposium on Multimedia Workshops 2007(ISMW’07)*, pp.10-12, 2007.
[9] S. Lee, G. L. Park: ”Analysis of the passenger pick-up pattern for taxi location recommendation”, in *IEEE 4th International Conference on Networked Computing and Advance Information Management*, 2008, pp.199-204.
[10] J. B. MacQueen: “Some methods for classification and analysis of multivariate observations,” in *Proceedings of the Fifth Symposium on Math, Statistics, and Probability*, 1967, pp.281-297.
[11] C. Spearman: "The proof and measurement of association between two things" *American Journal of Psychology*, vol.15, 1904, pp.72–101.

T. Takayama graduated University of Tsukuba in 1990. He received *Ph. D* from Tsukuba University in 1995. He worked Hiroshima City University from 1995 to 1998. From 1998 to 2002, he was an assistant professor of Iwate Prefectural University. From 2002, he is an associate professor of Iwate Prefectural University. His major includes database, data mining, and data engineering.

K. Matsumoto graduated Iwate Prefectural University in 2010. Currently, he works at IX Knowledge Inc.

A. Kumagai is currently a student of Iwate Prefectural University. She is engaged in the research of taxi information system.

N. Sato graduated Toyo University in 1999. He received *Ph. D* from Toyo University in 2004. He worked Toyo University from 2004 to 2007. From 2007, he is an assistant professor of Iwate Prefectural University. His major includes distributed database retrieval.

Y. Murata graduated University of Yamanashi in 1977. He received Master’s degree from Nagoya University in 1979. He worked at NTT and NTT Docomo Inc. from 1979 to 2006. He was engaged in development of mobile communication system. He received *Ph. D* from Shizuoka University in 2003. From 2006, he is a professor of Iwate Prefectural University. His major includes database application, mobile, and ubiquitous computing.