Using Probe Person Data for Travel Mode Detection

Muhammad Awais Shafique, Eiji Hato, Hideki Yaginuma

Abstract—Recently GPS data is used in a lot of studies to automatically reconstruct travel patterns for trip survey. The aim is to minimize the use of questionnaire surveys and travel diaries so as to reduce their negative effects. In this paper data acquired from GPS and accelerometer embedded in smart phones is utilized to predict the mode of transportation used by the phone carrier. For prediction, Support Vector Machine (SVM) and Adaptive boosting (AdaBoost) are employed. Moreover a unique method to improve the prediction results from these algorithms is also proposed. Results suggest that the prediction accuracy of AdaBoost after improvement is relatively better than the rest.

Keywords—Accelerometer, AdaBoost, GPS, Mode Prediction, Support vector Machine.

I. INTRODUCTION

DATA acquired from travel surveys provide basic information for the traffic modelling, service optimization and routing. This is immensely valuable for traffic planners, transportation authorities, public transport providers and researchers. Conventional data collection methods for travel surveys comprise telephone interviews, personal interviews, travel diaries, mail-back or web-based questionnaires, traffic counting on cross sections or intersections as well as analyses of transport schedule inquires. When collecting data on a large scale, most of these methods are expensive and time consuming. Consequently the update frequency of the data is very low. Moreover, nonresponse issues and underreported trips are well-known problems in surveys [1]-[3].

The second half of 1990s witnessed the introduction of Global Positioning System (GPS) devices to supplement the measurement of personal travel. One of the first household surveys employing GPS was conducted in 1997 in Austin, Texas, followed by many studies conducted to examine the application of GPS to determine travel behaviour [4]-[9]. These studies aim at determining the possibility of GPS device, combined with Global Information Systems (GIS), to replace or supplement conventional methods. Devices are either wearable or mounted in private vehicles. Common problems with GPS are loss of signal in underground facilities, high energy consumption and the willingness of users carrying the device.

Muhammad Awais Shafique is a Ph.D. candidate with the Department of Civil Engineering, University of Tokyo, Tokyo, Japan. (corresponding author; phone: +81-80-9448-1512; e-mail: awais@trip.t.u-tokyo.ac.jp).

Eiji Hato is a Professor at the Department of Civil Engineering, University of Tokyo, Tokyo, Japan (e-mail: hato@bin.t.u-tokyo.ac.jp).

Hideki Yaginuma is a Project Assistant Professor at the Department of Civil Engineering, University of Tokyo, Tokyo, Japan (e-mail: yaginuma@civil.t.u-tokyo.ac.jp).

II. RELATED WORK

The growing popularity of smartphones opens up novel opportunities for collecting data for travel surveys. Mobile phones have been studied in [10] for locating the positioning of the user for the purpose of tracking individual travel behaviour. Work reported in [11] proposes a method for assessing route choice models from smartphone GPS data, achieving reasonable results. Work presented in [12] proposes a method to automatically reconstruct trips in travel surveys using the data from the accelerometer sensor and GPS embedded in the smartphones.

Besides a GSM/UMTS module, modern smartphones are equipped with Assisted-GPS, wireless LAN and embedded sensors such as accelerometers, magnetometers or gyroscopes. This provides a high number of data sources, which can be intelligently engaged for identifying mobility behaviour.

This paper presents a novel approach to employ the embedded sensors of smartphones for detecting the individual mobility behaviour. For this purpose supervised learning algorithms like Support Vector Machines (SVM) and Adaptive boosting or AdaBoost are applied.

III. DATA COLLECTION

A. Cities Surveyed

This study comprises of the data collected from three Japanese cities namely Niigata, Gifu and Matsuyama. Fig. 1 shows the locations of the three cities with reference to the location of Tokyo.

Niigata is a coastal city facing the Sea of Japan and the Sado Island. It is the capital and the most populous city of Niigata prefecture. The city includes of a number of wetlands. In the west of Tokyo lies the Gifu prefecture, having a capital with the same name. The area varies from buildup city center to orchards in the surrounding regions.

Further south west of Gifu is the Ehime prefecture on the island of Shikoku. Its capital is Matsuyama. In addition to railway lines, Matsuyama also houses tram lines.

Population of each city (as reported in 2010 census) is given in Table I [13].

Niigata and Matsuyama both feature a humid subtropical climate. Gifu city experiences a wide range of weather throughout the year. The comparison of average temperature and precipitation during the survey months, among the three cities is summarized in Fig. 2 [14].

B. Data Description

The data collected can be classified into two categories, Location Data and Trip Data. Location data comprised of features extracted from accelerometer and GPS whereas trip data contained travel activity information. Table II shows the

contents of both types of data.

C. Data Collection Method

Location data was recorded using BCALs (Behavioral Context Addressable Loggers in the Shell) [15] or smart phones. BCALs are purpose-built devices equipped with accelerometer and GPS in addition to other sensors. Nowadays smart phones have both of these sensors integrated in them; therefore they can serve the same purpose as BCALs.

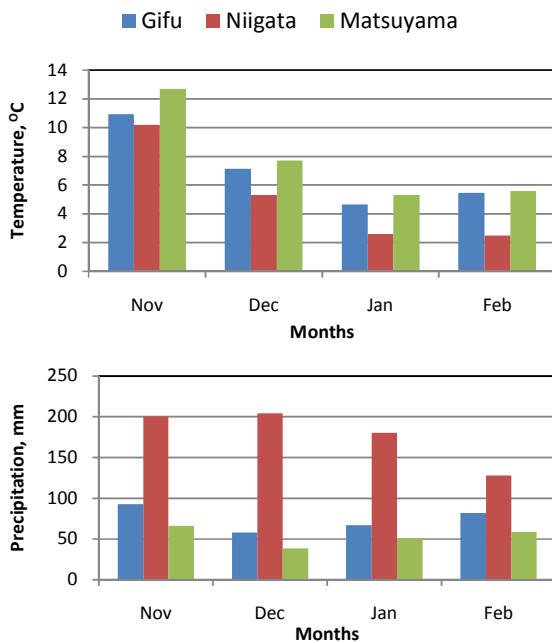Fig. 1 Location of Tokyo, Niigata, Gifu and Matsuyama

Fig. 2 Comparison of Temperature and Precipitation among cities

TABLE I
POPULATION OF CITIES SURVEYED

City	Population
Niigata	811,901
Gifu	413,136
Matsuyama	517,231

TABLE II
CONTENTS OF LOCATION AND TRIP DATA

Data	Contents
Location Data	User ID, Recording date and time, Exercise intensity, No. of steps, Min, Max and Avg. accelerations in movement, crosswise and vertical directions, Resultant and Avg. resultant accelerations
Trip Data	User ID, Departure date and time, Arrival date and time, Means of transportation

Trip data was logged using activity diaries or by the help of simple software in the smart phones. The respondents can select the mode of transportation used for the trip from within the software and start the trip. After reaching the destination the trip can be stopped and hence the activity can be recorded in the phone.

D. Amount of Data

The surveys were conducted during Jan to Feb 2011 in Niigata, Dec 2010 to Jan 2011 in Gifu and Nov 2010 to Jan 2011 in Matsuyama. Table III gives the amount of raw and processed data (discussed later in section V. Pre-Processing) used in this study. Moreover the distribution of processed data with respect to the mode of transportation is also shown in the table.

TABLE III
(A) AMOUNT OF RAW AND PROCESSED DATA

City	Raw Data	Processed Data
Niigata	341,712	229,502
Gifu	258,388	143,626
Matsuyama	507,014	269,375

(B) DISTRIBUTION BY MODE

Mode	Niigata	Gifu	Matsuyama
Walk	164,078 (71.49%)	83,645 (58.24%)	168,687 (62.62%)
Bicycle	3,214 (1.4%)	24,559 (17.09%)	15,404 (5.72%)
Car	61,785 (26.92%)	34,678 (24.14%)	81,653 (30.31%)
Train	425 (0.19%)	744 (0.52%)	3,631 (1.35%)

IV. ELEMENTARY ANALYSIS

A. GPS Data

Using GPS, the location of the respondent is recorded after every 1 minute. Hence the distance covered per minute can be assessed. This feature can help distinguish among different modes of transportation. Fig. 3 shows a part of the trip done by a respondent in Niigata city. The person walked from point A to F after which he used car from point F to K. The points are taken approximately 10 minutes apart to clarify the picture. It can be observed that the two modes can easily be separated based on the distance covered from point to point.

Fig. 3 Route formed by GPS data

B. Accelerometer Data

The accelerometer sensor integrated in the smart phone calculates and records the acceleration along X, Y and Z directions. These accelerations can also assist in determining the mode of transportation used. Fig. 4 shows the variation of average resultant acceleration along the same route (A-K).
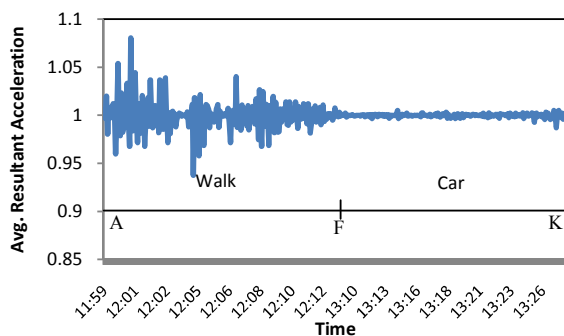

Fig. 4 Acceleration along the trip

It is evident from Fig. 4 that some pattern exists among the modes when it comes to acceleration.

Therefore this study used the patterns present within the GPS data and accelerometer data in order to determine and distinguish the modes.

V. PRE-PROCESSING

A. Mode Assignment

Before assigning the mode of transportation used, the location data for each city was scanned for any redundant entries, with reference to the trip data. Afterwards using the departure and arrival times listed in the trip data, the relevant data sets in the location data were assigned the corresponding mode of transportation. The data sets which were left unassigned were deleted.

B. Learning and Test Data

After assigning the mode of transportation to location data, it was used to generate learning data and test data.

Learning data was formed by randomly selecting some number of data sets for each mode. Random sampling with equal amount of data for each mode has resulted in better accuracy than complete random sampling. Due to the low amount of train data for Niigata city, the learning data was taken as 0.5% of the total data. For other cities approximately 1% of total data was used.

Test data contained all the data sets available, including the learning data. 13 features (mentioned in Table II) were selected to be used for prediction. In case of binary classifier, the learning data was reproduced into 4 files so that each file contained one distinct mode while the rest were renamed as "Other".

VI. METHODOLOGY

A. Support Vector Machine

The first choice for this type of classification problem was Support Vector Machine (SVM). SVM was trained by using the learning data for each city. After the training phase, the algorithm was used to predict the mode of transportation for the test data.

B. Adaptive Boosting

Next Adaptive Boosting or AdaBoost was employed. It was used in two ways as shown in Figs. 5 (a), (b).

Firstly AdaBoost was trained for one mode (as it is a binary classifier) and then used to predict the test data. For example the mode was "Train" then the prediction resulted in "Train" and "Other". Further the algorithm was trained for the second mode and prediction was done on the data sets predicted as "Other" from the first step. This procedure continued for all four modes. Consequently a small amount of test data predicted as "Other" remained in the end. SVM was used to predict this data.

In the second method, AdaBoost algorithm was trained for each mode separately but the prediction was done for the entire test data. The prediction results were analyzed and the data sets having only one mode predicted were assigned that mode. Rest all were assigned as "Other". The whole procedure was repeated for the remaining test data assigned as "Other". In the end SVM was used to finish off the task.
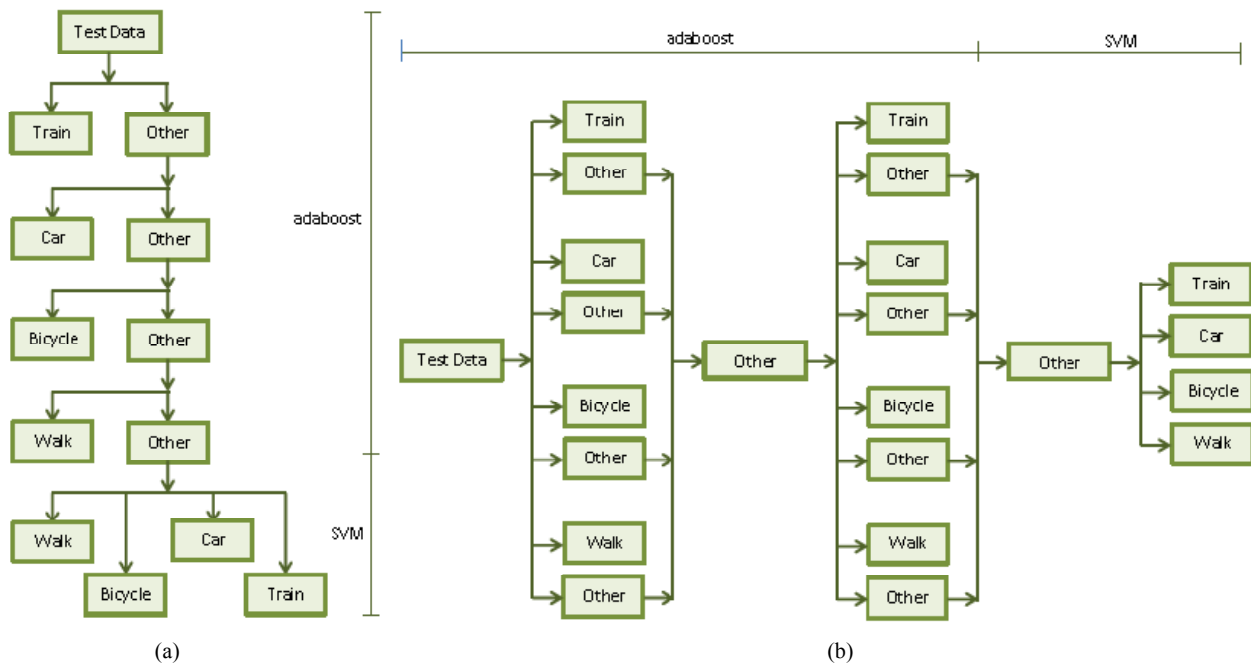
Fig. 5 (a) First method and (b) second method for AdaBoost application

C. Prediction Improvement

The prediction results, acquired by the application of SVM alone and by AdaBoost along with SVM, were further improved by adopting a simple method. By applying conditions on user ID, measurement date and measurement time, the data was divided into a number of trips. The mode is expected to remain same throughout one trip, so in light of this hypothesis the statistical mode of the results was determined for each trip and was then assigned to the respective trip.

VII. RESULTS & DISCUSSION

The prediction results acquired for the three cities, by applying the methodology discussed, are summarized in Table IV. The results suggest that the method of prediction improvement has a positive influence on the outcomes. Moreover the application of AdaBoost proved to be better than SVM alone. If the two methods of AdaBoost application are investigated then the second method involving repetition yielded comparatively better prediction results.

When cities are compared, Niigata gives the best results followed by Gifu and then Matsuyama. This might be because of difference in infrastructure, climate change and variation in technologies introduced. Although Niigata has the highest population among the three cities compared and the precipitation level is also very high, yet the prediction results are much better, partly due to moderate climate present during the survey.

VIII. CONCLUSION

From the results it is evident that determination of transportation mode is indeed very much possible by using GPS and accelerometer data. Although the prediction accuracy is different for each city yet this study can be considered as one step closer to the realization of the commercial use of smart phones in the transportation planning sector. Smart phones provide an immense opportunity to be utilized for trip reconstruction in travel surveys and studies like the one presented here, provide a means to achieve that goal.

The prediction accuracy is limited by the ambiguity of the features used in the prediction as well as the capacity of the algorithms employed. Therefore results can be further improved by including more features into the analysis and by introducing them to better learning algorithms.

TABLE IV
RESULTS OF PREDICTION ACCURACY

City	Mode	Normal			Improved		
		svm	AdaBoost (1)*	AdaBoost (2)*	svm	AdaBoost (1)	AdaBoost (2)
Niigata	Walk	81.77 %	81.76 %	82.38 %	92 %	91.39 %	92.49 %
	Bicycle	71.53 %	86.43 %	87.06 %	93.25 %	100 %	100 %
	Car	64.67 %	74.83 %	75.32 %	68.96 %	81.67 %	81.3 %
	Train	93.88 %	97.18 %	97.88 %	100 %	100 %	100 %
	All	77.05 %	79.99 %	80.57 %	85.83 %	88.91 %	89.6 %
Gifu	Walk	64.46 %	67.44 %	70.27 %	75.72 %	76.16 %	79.58 %
	Bicycle	72.09 %	73.89 %	74.13 %	86.05 %	90.76 %	89.83 %
	Car	68.65 %	75.84 %	76.28 %	75.64 %	83.76 %	86.23 %
	Train	87.5 %	96.91 %	97.04 %	84.95 %	84.95 %	84.95 %
	All	66.9 %	70.72 %	72.52 %	77.51 %	80.54 %	82.97 %
Matsuyama	Walk	67.75 %	68.65 %	68.43 %	79.54 %	79.99 %	79.76 %
	Bicycle	59.65 %	62.64 %	63.46 %	72.32 %	81.41 %	81.06 %
	Car	44.63 %	57.93 %	59.78 %	44.92 %	67.28 %	70.62 %
	Train	60.53 %	67.2 %	67.78 %	58.14 %	58.63 %	55.96 %
	All	60.18 %	65.04 %	65.51 %	68.34 %	75.93 %	76.74 %

*1 and 2 depicts the method used for the application of AdaBoost

REFERENCES

[1] W. Brög, E. Erl, A. H. Meyburg, & M. J. Wermuth, "Problems of Non-reported Trips in Survey of Non-home Activity Patterns." Transportation Research Record 891. 1982, pp. 1-5.
[2] A. Richardson, & E. Ampt, "Nonresponse Issues in Household Travel Surveys." Conference Proceedings 10 – Household Travel Surveys: New Concepts and Research Needs. National Academy Press, Washington D.C., USA, 1996, pp. 79-114.
[3] J. P. Zmud, & C. H. Arce, "Item Nonresponse in Travel Surveys: Causes and Solutions." Paper presented at the International Conference on Transport Survey Quality and Innovation, Grainau, Germany, 1997.
[4] E. Murakami, & D. P. Wagner, "Can using global positioning system (GPS) improve trip reporting?" Transportation Research Part C, 7, 1999, pp. 149-165.
[5] D. Pearson, "Global Positioning System (GPS) and Travel Surveys: Results from the 1997 Austin Household Survey." Paper presented at the 8th Conference on the Application of Transportation Planning Methods, Corpus Christi, Texas, 2001.
[6] J. Wolf, R. Guensler, & W. Bachman, "Elimination of the Travel Diary: An Experiment to Derive Trip Purpose From GPS Travel Data." Paper presented at the 80th Annual Meeting of the Transportation Research Board, January 7-11, 2001, Washington, D.C.
[7] P. Stopher, C. Fitzgerald, & J. Zhang, "Search for a Global Positioning System Device to Measure Person Travel." Transportation Research Part C, 16, 2007, pp. 350-369.
[8] W. Bohte, & K. Maat, "Deriving and validating trip purposes and travel modes for multi-day GPS-based travel surveys: A large-scale application in the Netherlands." Transportation Research Part C, 17, 2009, pp. 285-297.
[9] H. Gong, C. Chen, E. Bialostozky, & C. T. Lawson, "A GPS/GIS method for travel mode", 2012.
[10] Y. Asakura, & E. Hato, "Tracking survey for individual travel behavior using mobile communication instruments." Transportation Research Part C, 12, 2004, pp. 273-291.
[11] M. Bierlaire, J. Chen, & J. Newman, "Modeling Route Choice Behavior from Smartphone GPS data." Report TRANSP-OR 101016. Transport and Mobility Laboratory, Ecole Polytechnique Fédérale de Lausanne, October 2010.
[12] P. Nitsche, P. Widhalm, S. Breuss, & P. Maurer, "A strategy on how to utilize smartphones for automatically reconstructing trips in travel surveys." Procedia - Social and Behavioral Sciences 48, 2012, pp. 1033 – 1046.
[13] Statistics Bureau Japan website, Portal site of official statistics of Japan (accessed 16 July 2013). Available from: http://www.e-stat.go.jp/SG1/estat/NewListE.do?tid=000001039448.
[14] Japan Meteorological Agency website, Historic monthly values (accessed 15 July 2013). Available from: http://www.data.jma.go.jp/obd/stats/data/en/smp/index.html.
[15] E. Hato, "Development of behavioral context addressable loggers in the shell for travel-activity analysis." Transportation Research Part C 18, 2010, pp. 55–67.