RESEARCH ARTICLE OPEN ACCESS

Predicting The Mode Of Transportation Using GPS Data, For Vehicular Carbon Footprint Determination

Muhammed Awais Shafique , Eiji Hato
Department of Civil Engineering, University of Central Punjab, 1 Khayaban-e-Jinnah, Johar Town, Lahore 54000 Pakistan
Department of Civil Engineering, The University of Tokyo, 7-3-1, Hongo, Bunkyo-ku, Tokyo 113-8656, Japan
Corresponding Author: Muhammed Awais Shafique

ABSTRACT
Greenhouse gas emissions by vehicles is damaging the environment. In order to take remedial measures at individual level, one must first know the full scale of damage being done. This study suggests that using GPS data from smartphones, the travel mode used by individuals can be classified into motorized and non-motorized. It can assist in correctly estimating the vehicular carbon footprint by each individual. In this study, simplest features (travel duration, travel distance and average velocity), derived from GPS data, were used to train and test two popular algorithms i.e. Support Vector Machine and Random Forest. Results show that Random Forest provides a prediction accuracy of 90%, outperforming Support Vector Machine.
Keywords: Support Vector Machine, Random Forest, Machine Learning, Travel Mode, Carbon Footprint.

Date Of Submission:05-10-2018 Date Of Acceptance: 17-10-2018

I. INTRODUCTION

According to International Organization of Motor Vehicle Manufacturers, over the past thirty years, global car sales have more than doubled, from about 29 million in 1980 to 69 million in 2016. In Pakistan, car sales has seen a 25% increase in the last 5 years [1]. This ever-increasing car fleet is posing a threat to the environment due to the greenhouse gases emitted. In addition, large number of cars results in more traffic congestion and subsequently more CO_2 emissions enter the environment. Efforts to curb this threat start at individual level. We need to understand the damage we are doing to the environment by using personal vehicles for travelling. Gradually decreasing the number of car trips at individual level can have a compounding effect at a larger scale. To decrease the carbon footprint, we first need to have a clear and accurate estimation of our present state. When estimated ourselves, we usually have a low perception of damage done. Therefore, it is necessary that the greenhouse gas emissions be calculated passively for each individual so that an unbiased picture may be put forward.

For estimating the travel related gas emissions, smartphones can be very conveniently utilized as they are usually with the person making the trip. To determine the amount of vehicular miles and non-vehicular miles travelled, travel mode should be automatically predicted.

II. LITERATURE REVIEW

The increasing popularity and penetration of smartphones has opened up new opportunities in numerous fields. Mobile phones have been used to study the travel behavior of individuals by locating and tracking their positions [2]. GPS data collected by smartphones has also been utilized to assess route choice models [3]. Data from accelerometer and GPS has been used to reconstruct trips in travel surveys [4], to monitor busses [5].

GPS devices have been used by many researchers for the purpose of mode detection, whether employing rule-based algorithms [6-9], or machine learning algorithms [10-13]. Employing machine learning algorithms, this paper uses the simplest features i.e. trip duration, trip distance and average velocity, derived from GPS, to predict whether the user is using motorized mode (vehicular greenhouse gas emitted) or non-motorized mode (no emission of vehicular greenhouse gases). This study is first of its kind, undertaken in Pakistan.

III. METHODOLOGY

3.1 Data Collection

GPS data was collected by nine students from University of Engineering and Technology (UET), Lahore, using smartphones over a period of about 45 days (Oct. 2016 to mid Nov. 2016). An android application named “PP” was used by the users to report the start and end of each trip as well as the mode used for each trip. The data recorded for each trip contained the travel time in seconds, travel distance in meters and average velocity in km/hr. Trips with missing or improper data were screened and removed. Table 1 summarizes the number of trips recorded for motorized (Bike, Rickshaw, Car, Bus) and non-motorized (Walk, Bicycle) modes.

For training the classification algorithms, 75% of data was sampled using stratified random sampling. The remaining 25% was used to test the algorithm to check its training capability and prediction accuracy.

Table 1 – Amount of Data Collected

Mode		No. of Trips	
Non-motorized	Walk	54	58
	Bicycle	4	
Motorized	Bike	73	141
	Rickshaw	16	
	Car	25	
	Bus	27	
Total	199	199	

3.2 Classification Algorithm

Two popular algorithms were used and compared for classifying the travel mode data.

3.2.1 Support Vector Machine

Support Vector Machine (SVM) is a state-of-the-art classification method, which has proved to be immensely useful in a wide range of fields including robotics, text recognition, bioinformatics and image recognition [14]. SVM uses kernels to solve non-linear problems by linear methods. Three types of kernels were studied in this paper i.e. linear, radial or polynomial. The associated parameters were varied, as shown in Table 2, to determine the values that maximized the prediction accuracy.

3.2.2 Random Forest

Random forest is an ensemble classification and regression method. At the training level, it constructs a number of decision trees. Each decision tree is then used to predict the class (in this case mode). All predictions from individual trees are used to conclude the final prediction based on maximum votes. Unlike Decision Trees, Random Forest does not require pruning. The number of trees and maximum variables for split were varied as demonstrated in Table 2.

Table 2 – Range of Parameter Values for SVM and RF

Parameters	SVM			Random Forest
	Linear	Radial	Polynomial	
gamma	-	0.01 - 10	0.01 - 10	-
degree	-	-	1 - 5	-
coef0	-	-	0 - 1	-
no. of trees	-	-	-	100 - 300
max. variables	-	-	-	1 - 3

IV. RESULTS AND DISCUSSION

The comparison of parameters provided with the most suitable values for our data, provided in Table 3. Using these parameter values, the algorithms were trained, following which the test data was fed to determine the prediction accuracy. Results, provided in Table 4, establishes that Random Forest performs better than Support Vector Machine, accurately predicting 90% of the trips. In case of SVM, polynomial kernel outperforms both linear and radial kernels. The misclassifications can be the result of algorithm confusing bike and rickshaw with non-motorized modes

V. CONCLUSION AND FUTURE WORK

This study shows that GPS data collected in Pakistan can be used successfully to classify between motorized and non-motorized modes. Between the two classifiers compared, Random Forest gives better prediction accuracy. Once the motorized and non-motorized modes are classified, the vehicular greenhouse gas emissions can be estimated much more accurately. This will show an unbiased picture to the individuals about their contribution towards deteriorating the environment and thus will urge them to change their travel patterns and preferences in order to decrease their impacts. This is an ongoing research so currently the data is small. The results might improve once more data is available to better train the algorithm. Individual modes will also be predicted in future.

Table 3 – Suitable Parameter Values for SVM and RF

Parameters	SVM			Random Forest
	Linear	Radial	Polynomial	
gamma	-	0.5	10	-
degree	-	-	4	-
coef0	-	-	1	-
no. of trees	-	-	-	100
max. variables	-	-	-	2

Table 4 – Prediction Results for SVM and RF

Mode	No. of Trips			No. of Test Trips correctly predicted, (%)			
				SVM			RF
	Total	Train	Test	Linear	Radial	Polynomial	
Motorized	141	106	35	28	28	30	32
Non-motorized	58	44	14	13	13	13	12
Total	199	150	49	41 (84)	41 (84)	43 (88)	44 (90)

REFERENCES

[1]. OICA. World motor vehicle sales. 2017. http://www.oica.net/category/sales-statistics/ [accessed 15.05.17].

[2]. Asakura Y, Hato E. Tracking survey for individual travel behaviour using mobile communication instruments. Transportation Research Part C: Emerging Technologies. 2004 Aug 31;12(3):273-91.

[3]. 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

[4]. 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.

[5]. Osama A., Mahdy H., Kandil K., & Elhabiby M. (2016). Monitoring the impact of unscheduled stops and unscheduled route deviations on bus service using Automatic Vehicle Location system: a case study. Advances in Transport Studies: an international journal. 39; 37-48.

[6]. Stopher P, FitzGerald C, Zhang J. Search for a global positioning system device to measure person travel. Transportation Research Part C: Emerging Technologies. 2008 Jun 30;16(3):350-69.

[7]. Bohte W, Maat K. 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: Emerging Technologies. 2009 Jun 30;17(3):285-97.

[8]. Chen C, Gong H, Lawson C, Bialostozky E. Evaluating the feasibility of a passive travel survey collection in a complex urban environment: Lessons learned from the New York City case study. Transportation Research Part A: Policy and Practice. 2010 Dec 31;44(10):830-40.

[9]. Gong H, Chen C, Bialostozky E, Lawson CT. A GPS/GIS method for travel mode detection in New York City. Computers, Environment and Urban Systems. 2012 Mar 31;36(2):131-9.

[10]. Bolbol A, Cheng T, Tsapakis I, Haworth J. Inferring hybrid transportation modes from sparse GPS data using a moving window SVM classification. Computers, Environment and Urban Systems. 2012 Nov 30;36(6):526-37.

[11]. Ellis K, Godbole S, Marshall S, Lanckriet G, Staudenmayer J, Kerr J. Identifying active travel behaviors in challenging environments using GPS, accelerometers, and machine learning algorithms. Public Health. 2014 Oct 23;2:39-46.

[12]. Shafique MA, Hato E. Use of acceleration data for transportation mode prediction. Transportation. 2015 Jan 1;42(1):163-88.

[13]. Shafique MA, Hato E. Travel mode detection with varying smartphone data collection frequencies. Sensors. 2016 May 18;16(5):716.

[14]. Boser BE, Guyon IM, Vapnik VN. A training algorithm for optimal margin classifiers. InProceedings of the fifth annual workshop on Computational learning theory 1992 Jul 1 (pp. 144-152). ACM.

Muhammed Awais Shafique "Predicting The Mode Of Transportation Using Gps Data, For Vehicular Carbon Footprint Determination "'International Journal of Engineering Research and Applications (IJERA) , vol. 8, no.10, 2018, pp 70-72