GEPE Group of Energy and Power Electronics

CENTROALGORITMI

Universidade do Minho

João C. Ferreira, Vítor Monteiro, José A. Afonso, João L. Afonso

"Tracking Users Mobility Patterns Towards CO2 Footprint"

Advances in Intelligent and Soft Computing, 1st ed., Sigeru Omatu, Ali Selamat, Grzegorz Bocewicz, Pawel Sitek, Izabela Nielsen, Julián A. García-García, Javier Bajo, Ed. AISC Springer Verlag, 2016, Part I, pp.87-96.

http://link.springer.com/chapter/10.1007%2F978-3-319-40162-1_10

ISBN: 978-3-319-40161-4 (Print) 978-3-319-40162-1 (Online)

ISSN: 2194-5357

DOI: 10.1007/978-3-319-40162-1_10

João C. Ferreira, Vítor Monteiro, José A. Afonso, João L. Afonso. "***Tracking Users Mobility Patterns Towards CO2 Footprint***", in Advances in Intelligent and Soft Computing, 1st ed., Sigeru Omatu, Ali Selamat, Grzegorz Bocewicz, Pawel Sitek, Izabela Nielsen, Julián A. García-García, Javier Bajo, Ed. AISC Springer Verlag, 2016, Part I, pp.87-96. ISSN: 2194-5357 DOI: 10.1007/978-3-319-40162-1_10 ISBN: 978-3-319-40161-4 (Print) 978-3-319-40162-1 (Online)

Tracking Users Mobility Patterns Towards CO_2 Footprint

João C. Ferreira[1,3], Vítor Monteiro[1], José A. Afonso[2], João L. Afonso[1]

[1]Centro ALGORITMI, University of Minho, Guimarães, Portugal,
[2]CMEMS-UMinho, Guimarães, Portugal, and [3]ADEETC at ISEL, Lisbon, Portugal

Abstract. This research work is based on the development of a mobile application and associated central services for tracking users' movements in a city, identifying the transportation mode and routes performed. This passive tracking generates useful data about users' habits, which are then associated with the CO_2 emission in the form of a mobility invoice, with the goal of enabling the users to understand their carbon footprint resulting from the users' mobility process in the city. The performance of the developed system is validated through experimental tests based on data collected during six months from more than 2500 mobility experiences.

Keywords: mobile application, personalized data, geographical system, intelligent public transportation, carbon footprint.

1 Introduction

CO_2 emissions in big cities due to transportation systems raise the need to improve the sustainability and accessibility of collective transport, while simultaneously promoting the use of more environmentally friendly transportation systems. In this sense, it is important to make available adequate and updated information regarding the mobility options offered by transport operators and users. One important research work is to create a tool to measure the mobility of people in a city, identifying passively the transportation mode, routes performed and associated times. Information and communication technologies (ICT) have the potential to effectively change the way people live and their mobility and energy consumption. Nowadays, mobile devices incorporates many diverse and powerful sensors, like GPS, cameras, microphones, light, temperature, direction (i.e., magnetic compasses) and acceleration (i.e., accelerometers). Accelerometers with GPS can be used to perform activity recognition [1], a task which involves identifying the physical activity of a user. Activity inference provides the 'what' of a user's context, whereas location sensors (such as cell-tower/Wi-Fi localization and/or GPS) provide the 'where'. This 'what' and 'where' information can be used by a number of mobile phone applications, including physical fitness and health monitoring [2], recommendation systems, and the study of environment and personal behavior. In this work, we explore the potential of a mobile device application designed to track users' habits in a customized way, using integrated accelerometers and GPS information, with the goal to create a monthly user invoice related with their movements and used transportation modes (e.g., bus, train, car, bike, walk, etc.). This approach allows the discovery of mobility habits of millions of users passively, just by having them carry mobile phones in their pockets. From the collected sensor data, it is possible to identify the user's transportation mode and also characterize the traffic conditions. In association with the travelled distance, we can track the CO_2 emissions resulting from the user mobility and provide the information to the user in the form of a monthly invoice related with the concept of carbon footprint [3].

The rest of the paper is organized as follows. Section 2 describes the tasks associated to data acquisition and classification. Section 3 describes the developed transportation mode identification system and the corresponding performance evaluation results. Section 4 concerns to the user mobility patterns, while Section 5 presents the proposed mobility invoice system. Section 6 presents the mobility advisor to reduce the carbon footprint. Finally, Section 5 presents the conclusions and future work.

2 Data Acquisition and Classification

The applied methodology is composed by a set of sequential processes. The first process, identified as data collection, varies from case to case, being responsible for the collection of huge amounts of data (big data). For testing purposes, we used mobile device sensor data from 50 Lisbon area users, in a period corresponding to the first six months of 2013. This phase involves the identification of outliers and the removal of inconsistent data to reduce the number of records [4]. The second process consists on data transformation into predefined classes (Fig. 1). This is a process that is specific to each case study. Taking into account the mobile device sensor data, Fig. 1 shows the 20 predefined classes (C_1 to C_{20}). These classes are created based on accelerometer measurements in the three orientation axis, as well as additional GPS data. The accelerometer data is divided into three dimensions: Z (vertical) is the upright direction; Y (longitudinal) is the direction of movement and X (transversal) is the steering direction. The transformation process to merge original data into these predefined classes uses information of two consecutive accelerometer data measurements, where the data value difference is classified into four scales: 1 - straight, when the measurement changes in the module are less than 0.1 ms^{-2}; 2 - smooth, when it is more than 0.1 ms^{-2} and less than 0.5 ms^{-2}; 3 - rough, when it is between 0.5 ms^{-2} and 2 ms^{-2}; 4 - very rough, when it is more than 2 ms^{-2}.

3 Transportation Mode Identification

From the mobile device sensor data, it is possible to identify the transportation mode that the user takes to go from A to B. Transportation mode detection has been explored by [5-7], among others. All of these approaches use past data to build a classification model that identifies the transportation mode and most of these approaches use a combination of GPS and accelerometer data from the three axes. From this data, it is possible to calculate the speed and the position. We developed this work based on a discrete

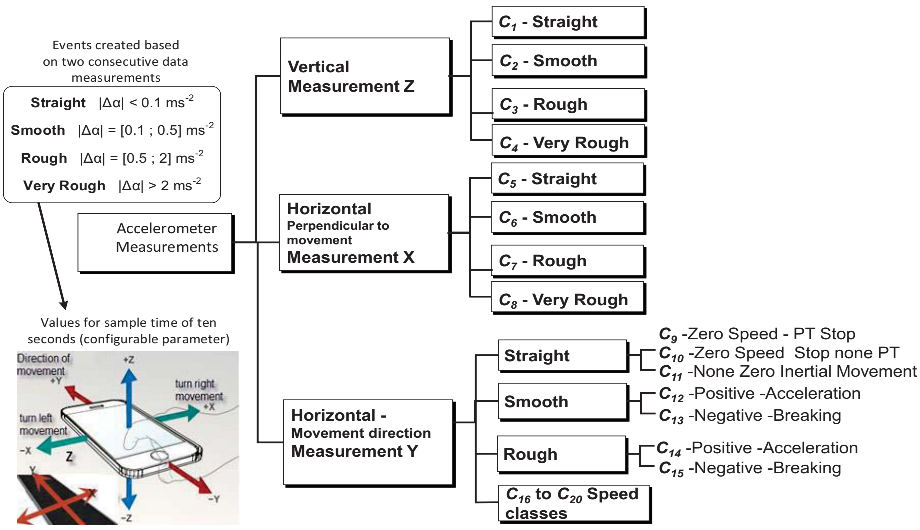

Fig. 1. Predefined data classes.

approach using predefined data classes and a training set of 250 cases, representing car travel (60), bus (50), train (30), underground (40), walking (25), boat (20) and motorcycle (25). Table 1 shows an example of event count per class for three transportation modes (other modes are omitted for simplicity purposes). Since we want a generic approach, the main effort to perform is the transformation of raw data into these predefined classes, as Table 1 shows. These classes can be increased to cover new situations, and when not used should be treated as empty fields.

Our approach for public transportation mode identification is based on the value of $P(TM_j|C_k)$, which means the probability of transportation mode j (TM_j) given C_k measurements are discrete classes defined from accelerometer data as illustrated on Figure 1. We use classes C_1 to C_{20}, using the Bayes theorem:

$$P(TM_j|C_k) = \frac{P(C_k|TM_j)}{P(C_k)} \propto P(C_k|TM_j)P(TM_j) \qquad (1)$$

where, TM_1 = Car, TM_2 = Bus, TM_3 = Train, TM_4 = Metro, TM_5= Walk, TM_6 = Boat and TM_7 = Motorcycle. We calculate $P(TMj)$ as the number of cases for transportation mode j divided by the number of cases in the training set, for example, $P(TM_1)$ = 60/250 = 0.24. The same approach is used to calculate the other values. For the probability $P(C_k|TMj) = P(\{C_1,C_2,..,C_{20}\}|TMj)$, we assume the independence of events, therefore:

$$P(C_k|TM_j) = \prod_{k=1}^{20} P(C_k|TM_j)\, P(TM_j) \qquad (2)$$

where, $P(C_k|TMi)$ is based on the training data set. Since the number of events varies with the sampling time and the route distance, we perform a normalization using the percentage. For example, for X-axis accelerometer data from car samples, we have 23 C_1 events, 2111 C_2 events, 3312 C_3 events and 34 C_4 events. This totals 5480 events, so we have 0.4% of events in class C_1, 38.5% in class C_2, 60.5% in class C_3 and 0.6% in class C_4. To avoid zero probability, since (2) is a multiplication, we always introduce

Table 1. Sample of data training set with three transportation modes. All values represent number of events after the second process.

Data Class	Car (12 km)	Bus (7.5 km)	Metro (2.3 km)
C_1	23	314	2313
C_2	2111	1350	112
C_3	3312	4175	2
C_4	34	65	0
C_5	3402	3603	967
C_6	1501	2421	1223
C_7	425	165	121
C_8	126	6	0
C_9	9	454	367
C_{10}	390	420	8
C_{11}	1234	1242	634
C_{12}	1476	1580	1021
C_{13}	1021	1374	32
C_{14}	521	412	53
C_{15}	631	556	41
C_{16}	254	2860	430
C_{17}	532	642	267
C_{18}	264	1210	412
C_{19}	1375	1340	810
C_{20}	96	0	0

an offset of one in the counting process. This is a similar process used in text classification through NB [8]. Taking into consideration that the 60 cases correspond to trips performed by car, in the calculation of the average of these values we reached the value of 0.5% for $P(C_1|TM_1)$. From other values of the training set we have $P(C_1|TM_2) = 10\%$ (this is higher because of the waiting time at bus stations, where the user is immobile, which means that C_1 events are being collected), $P(C_1|TM_3) = 95\%$, $P(C_1|TM_4) = 91\%$, $P(C_1|TM_5) = 15\%$, $P(C_1|TM_6) = 5\%$, and $P(C_1|TM_7) = 2\%$. This collection process with the class identification allows the determination of the transportation mode because some events are characteristic of some transportation modes in particular. For example, events in the C_1 class are common in railroad or underground transportation, which implies very smooth changes in Z-axis acceleration. On the other hand, if we have the majority of events in class C_4, this indicates a motorcycle or a boat. Meanwhile, the study of the acceleration shape allows differentiating the boat from the motorcycle. C_3 means several speed limitation bumps or potholes in the road in a vehicle or bus. Again, through the pattern of acceleration, it is possible to distinguish a pothole from a speed bump. Acceleration data from the X-axis (C_8 and C_7 classes) can be helpful to identify aggressive or drunk drivers.

Speed information from GPS is used to differentiate among walking, bicycling, boat, and other transportation modes. Periodic stops are used to differentiate among car/motorcycle and underground, bus or train. Motorcycle is better discerned from car transportation in high traffic periods, because the average speed is higher and the position pattern is different. In order to distinguish metro from train, we use the following heuristics: (1) Underground usually runs below ground without a GPS signal; (2) Distances

Table 2. Confusion matrix for the transportation mode identification.

	TM_1	TM_2	TM_3	TM_4	TM_5	TM_6	TM_7
TM_1	95%	3%	1%	0%	0%	0%	1%
TM_2	3%	89%	3%	2%	1%	0%	2%
TM_3	1%	3%	70%	24%	1%	0%	1%
TM_4	0%	2%	24%	72%	1%	1%	0%
TM_5	0%	1%	1%	1%	97%	0%	0%
TM_6	0%	0%	0%	1%	0%	82%	17%
TM_7	1%	2%	1%	0%	0%	17%	79%

between stops in underground transportation are usually smaller; (3) Altimetry information. Given the sensor information and GPS traces, we predict the transportation mode among the available modes. This is done calculating the probability $P(TM_j|D_k)$, with ($j = 1,\ldots 7$) and ($k = 1,\ldots 20$), and choosing the highest value. Table 2 presents performance results in a confusion matrix. Results are available using 6 month data in more than 2500 recorded mobility experiences. From this data we used 250 for training purposes and evaluated 500 cases based on precision measurement (number of correct cases classified for that transportation mode divided by the transportation mode cases available). We achieved high precision identification values for walking (97%) and car (95%), having lower values for train (70%) and underground (72%). In 24% of cases there was a classification change from underground to train.

4 User Mobility Patterns

Other knowledge that can be extracted from the collected mobile device sensor data is related with the information of the distinct locations where the users spend their time throughout the day (e.g., home, work, shopping centers, restaurants, etc.). From the data that we have collected, we are particularly interested in identifying locations where people spend a great deal of time, and associating these locations with in-formation about the environment, obtained from geographic information system data sources. We have all GPS data and time stored in a user mobility profile, in a cloud database, with the information about time and routes (XML graph with time and GPS coordinates). It is possible to present the route representation for that month with associated information of the transportation mode, the number of times the route was performed, and also the temporal periods. Thus, it is possible to represent the time that a user spent in each location. User mobility profile is stored in central server. Based on the information shown in Fig 2, it is possible to produce a monthly invoice.

5 Mobility Invoice System

Presently, carbon footprint is the most popular measure of environmental impact, and it is used to refer the amount of greenhouse gas emissions that are produced during the mobility process of a user. Using the knowledge discovery of transportation mode de-

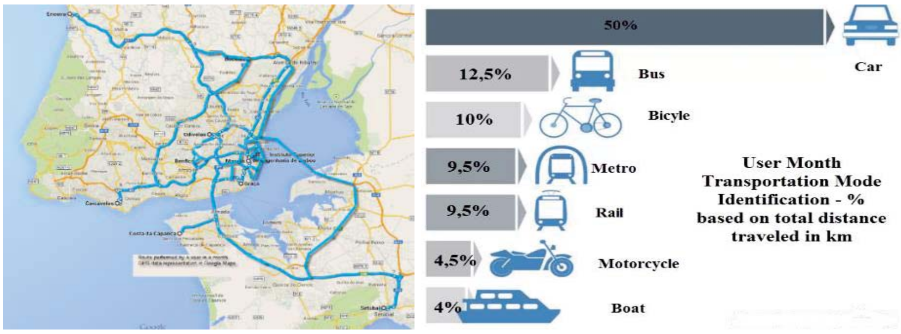

Fig. 2. User movement in a monthly period and associated transportation mode.

Mobility Invoice

Month: March 2014

Client nº: 1234

	km	Price per km	Total
Car	48	0.10	4.80 €
Public Transportation	135	0.01	1.35 €
Carpooling	102	0.03	3.06 €
			9.21 €

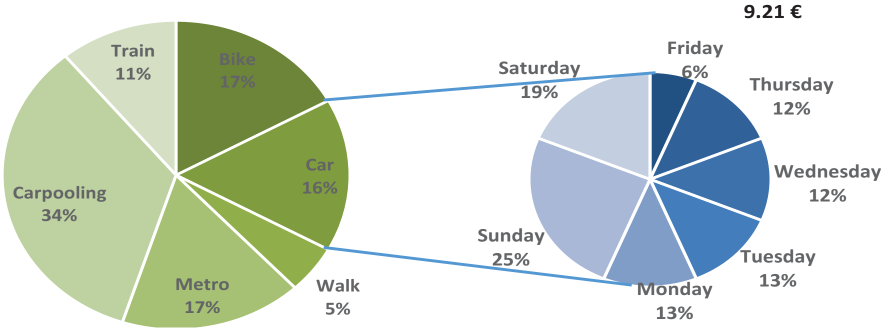

Fig. 3. Example of a mobility invoice, based on transportation mode identification and associated distance.

tection and associated distances, we are able to generate a monthly report with the associated CO_2 emissions (see Fig. 3), where a price per km is defined for each transportation mode [9][10]. From the tracking of user movements, it is possible to improve public transportation routes and timing. Our intention is to show an important output, which is a result of the passive data collection from mobile devices, and the consequent application of knowledge discovery to sustainable transportation. It is possible to develop a suggestion system to reduce this invoice and, consequently, reduce the CO_2 emissions in the city. This reduction can be achieved based on: (1) Carpooling system to reduce car usage; (2) Public transportation information integration [11]. For this purpose, a mobile application, developed based on current user position and taking into account the user habits (past route performed), provides personalized advice to go from current position to final destination based on public transportation availability (user

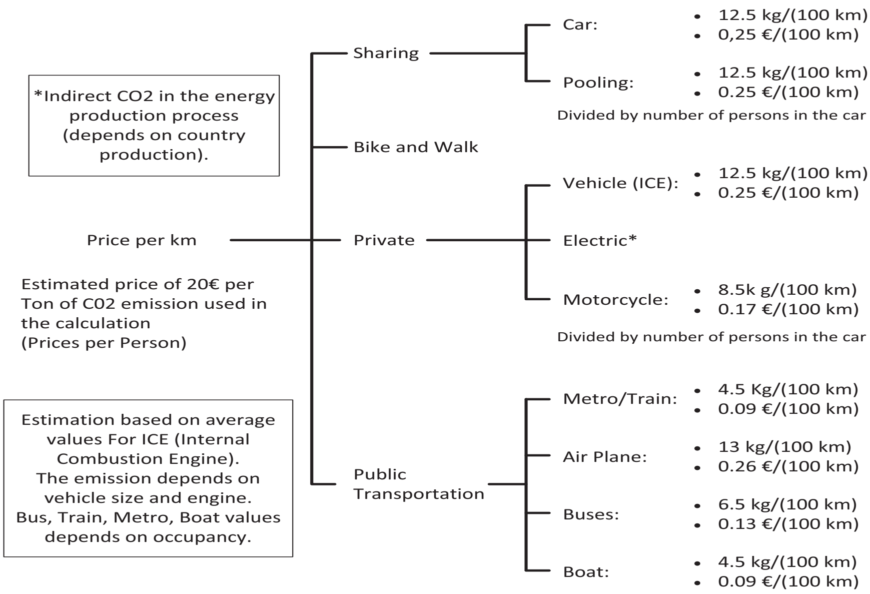

Fig. 4. Identification of the price per km of CO_2 emissions based on transportation mode.

receives route and scheduling information). An example of a mobility invoice is showed in Fig 3, where users travel distance by transportation mode are used to get a value, related with CO_2 emissions price based on values identified in Fig 4.

6 Mobility Advisor to Reduce the Carbon Footprint

The main idea is to provide advice to reduce the carbon footprint based on historical user mobility data, through the matching of this data with public transport services using the information of available routes and schedules. This approach can be complemented with data integration of different information sources, such as: car and bike sharing systems [11][12] and also carpooling [13]. These systems are important to the growing interest of the sustainable transport systems, and to reduce the growth of energy use, noise, air pollution and traffic congestion, due to a decrease in the number of cars and to an increase in the shared use of electric or low pollutant vehicles. Also, bike sharing systems permit door-to-door ride features and allow to access areas of the city that are forbidden for other kind of vehicles. Also, the integration of public transportation data can be used as complement. Through previous work, we integrate the access to several public transportation databases using a semantic approach described in [13] and the START project [http://www.start-project.eu], where the integrated information available in mobile devices increases the usage of public transportation among different experiences in European project partners (England, France, Spain and Portugal). With this, it is possible to have information about the schedule and the price of public transportation. With this information in a previous project we build a route planner that integrates routes and public transportation [11].

7 Conclusions

Tracking of user activity and associated mobility is now possible at low costs through the use of mobile device sensor information. The data generated using this approach is in the class of big data and has great impact in the study of user mobility habits. . In this work we show the application of this approach to the tracking of user mobility in a city through the identification of the used transportation modes, routes and times. This information can be transformed in an informative CO_2 invoice, which allows users to be aware of their carbon footprint. Along with that, associated measures and suggestions to reduce this invoice are then provided. With a considerable number of users, this passive tracking of data from citizens could generate useful data about mobility habits and be used to improve the citizens' mobility. Moreover, public transport operators can use the processed data to improve their transportation offers towards the users' needs.

8 References

1. Wazir Z. Khan, Yang Xiang, Mohammed Y Aalsalem, Quratulain Arshad, "Mobile phone sensing systems: A survey," Comm. Surveys & Tutorials, vol.15, pp.402-427, Feb. 2013.
2. Alexandros Pantelopoulos, Nikolaos G. Bourbakis, "A survey on wearable sensor-based systems for health monitoring and prognosis," IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews, vol.40, no.1, pp.1-12, Jan. 2010.
3. Patrick Luckow, E. A. Stanton, Bruce Biewald, Jeremy Fisher, F. Ackerman, E. Hausman, "2013 Carbon Dioxide Price Forecast," Synapse – Energy Economics, Inc., Nov. 2013.
4. João Ferreira, Jose Almeida, Alberto Silva, "The Impact of Driving Styles on Fuel Consumption: A Data Warehouse and Data Mining based Discovery Process," IEEE Transactions on Intelligent Transportation Systems, vol.16, no.5, pp.2653-2662, Mar. 2015.
5. S. Reddy, M. Mun, J. Burke, D. Estrin, M. Hansen, M. Srivastava, "Using Mobile Phones to Determine Transportation Modes," ACM Trans. on Sensor Networks, v6, n2,13-40, 2010.
6. D. J. Patterson, Lin Liao, D. Fox, and H. Kautz, "Inferring High-Level Behavior from Low-Level Sensors," UbiComp 2003: Ubiquitous Computing, vol.2864, pp.73-89, Oct. 2003.
7. Yu Zheng, Quannan Li, Yukun Chen, Xing Xie, Wei-Ying Ma, "Understanding mobility based on GPS data," International Conf. on Ubiquitous Computing, pp.312-321, Sept. 2008.
8. A. McCallum, K. Nigam, "A comparison of event models for naive bayes text classification," AAAI Press, Workshop on Learning for Text Categorization, vol.752, pp.41-48, 1998.
9. P.Luckow, Elizabeth A. Stanton, B. Biewald, Jeremy Fisher, Frank Ackerman, Ezra Hausman, "2013 Carbon Dioxide Price Forecast," Synapse – Energy Economics, Inc., Nov. 2013.
10. F. Ackerman, E. A. Stanton, "Climate Risks and Carbon Prices: Revising the Social Cost of Carbon," Economics: The Open-Access, Open-Assessment E-Journal,v6,pp.1-25,Apr. 2012
11. João C. Ferreira, A. R. Silva, João L. Afonso, "EV-Cockpit – Mobile Personal Travel Assistance for Electric Vehicles," Microsystems for Automotive Applications. Smart Systems for Electric, Safe and Networked Mobility. Springer, 2011.
12. João C. Ferreira, Vítor Monteiro, José A. Afonso, João L. Afonso, "Mobile Cockpit System for Enhanced Electric Bicycle Use," IEEE TII, vol.11, no.5, pp.1017-1027, Oct. 2015.
13. João C. Ferreira, Paulo Trigo, Porfirio Filipe, "Collaborative Car Pooling System," International Conf. on Sustainable Urban Transport and Environment, Paris, pp.1-5, June 2009.
14. João C. Ferreira, Porfirio Filipe, Paulo Matos Martins, "Cooperative Transportation Infrastructure," ICITIS Int. Conf. on Civil, Transport and Env. Eng., Malaysia, pp.1-5, Aug. 2013.