Transportation Research Part C 71 (2016) 434–446

Contents lists available at ScienceDirect

Transportation Research Part C

journal homepage: www.elsevier.com/locate/trc

Crowdsourcing mobility insights – Reflection of attitude based segments on high resolution mobility behaviour data ☆

Ivana Semanjski *, Sidharta Gautama

Ghent University, St-Pietersnieuwstraat 41, B-9000 Gent, Belgium

ARTICLE INFO

Article history:
Received 31 July 2015
Received in revised form 27 May 2016
Accepted 23 August 2016
Available online 30 August 2016

Keywords:
Data-driven mobility management
Attitude based segmentation techniques
Crowdsourcing mobility behaviour
Smartphone based mobility data collection
Support vector machines
Smart city mobility

ABSTRACT

Recently, the use of market segmentation techniques to promote sustainable transport has significantly increased. Populations are segmented into meaningful groups that share similar attitudes and preferences. This segmentation provides valuable information about how policy options, such as pricing measures or advertising campaigns, should be designed and promoted in order to successfully target different user groups. In this paper, we aim to bridge between psychological, social marketing and ICT research in the field of transportation. We explore how attitude based segments are reflected in high resolution mobility behaviour data, crowdsourced via mobile phones. We use support vector machines to map eight attitudinal segments, as defined under the European project SEGMENT, to the n dimensional space defined by crowdsourced data. The success rate of the proposed approach is 98.9%. This demonstrates the applicability of the method as a way to automatically map attitudinal segments to a wider population based on observed mobility data instead of using explicit attitudinal surveys. In addition, the proposed approach can facilitate the delivery of personalised target messages to individuals (e.g. via smartphones) or at target locations where users, belonging to specific segment, are located at specific time windows since the data includes the time-space indications.

1. Introduction

Based on the International energy agency findings (IEA, 2014), transportation contributes about 25% to the global CO_2 emission and is the only major sector where emissions continue to grow. Even though technological advances improved the energy efficiency in transport, nevertheless this has been outweighed by the increase in travel demand. In recent years, several strategies have been explored in order to lower the demand and facilitate the users' shift towards the more sustainable means of transportation. Almost all of these strategies rely on the concept of target groups, where the complexity and heterogeneity of the whole population is reduced by dividing it into relevant subgroups for which specific mobility management campaigns and policies are developed. And whether studies focus on just one target group (Bamberg et al., 2007; Delang and Cheng, 2012; Hu et al., 2013) or analyse the whole population (Diana and Pronello, 2010; Prillwitz and Barr, 2011), the impact of traditional market segmentation techniques in the field of mobility is evident. The first applications of these techniques were based on the socio-demographic segmentation and have shown that age, gender, occupation, household size, income and car ownership are highly relevant for mobility behaviour (De Jong et al., 2004). Next to the

☆ This article belongs to the Virtual Special Issue on: Data-driven smart-city-enabled traffic system modeling, analysis, and optimization.

* Corresponding author.
E-mail addresses: ivana.semanjski@ugent.be (I. Semanjski), sidharta.gautama@ugent.be (S. Gautama).

http://dx.doi.org/10.1016/j.trc.2016.08.016

socio-demographic segmentation, a behavioural segmentation was also used to define segments based on the usage of different transportation modes and frequency of their use (Prillwitz and Barr, 2011). In addition, segmentation by lifestyle was adopted to better describe an individual's daily range of actions (Redmond, 2000), but none of these successfully explained the underlying individual values systems and attitudes that are most likely to have an impact on willingness to change mobility behaviour and adopt more sustainable means of transportation. Therefore, the most recent research adopts an attitude based segmentation approach in order to better understand users' motivation. The attitude based segmentation approach relies on rational choice models like the theory of planned behaviour (Ajzen, 1991). Here, next to social norm and perceived behavioural control, person's intention to perform a behavioural option is causally determined by the attitude towards the behavioural option (individual's general feeling of favourableness or unfavourableness towards this option). Attitude mirrors the beliefs that a person holds about the positive and negative consequences of mobility behaviour and the values that this person ascribes to those consequences. Therefore, attitude based segments are subgroups of population that share similar attitudes towards same behavioural options and have proven to have a high predictive power for transportation mode selection (Hunecke et al., 2010; Eriksson and Forward, 2011). The related studies, similar to the above segmentation approaches, are based on surveys over the selected sample of participants for data collection (Bamberg et al., 2007; Anable, 2005; Anable and Wright, 2013).

Nevertheless, today's advances in data collection (primary high resolution behaviour data crowdsourced from active mobile devices) (Cheng et al., 2012; Patire et al., 2015; Wu and Liu, 2014) and processing (Gong et al., 2012; Goodchild and Li, 2012; Calabrese et al., 2013) provide new insights into mobility behaviour far more detailed than the one collected via traditional travel diaries, household surveys or surveys in general. To get an impression of the difference in resolution of collected data, one can compare a typical travel survey input on how often, by which mode and what distance a person travels to work (e.g. five times a week; by public transportation; around 10 km) with crowdsourced time-space information and automatic transportation mode detection (where we have information on which public transit stops were used; how did (s)he reach the public transportation network; at what time and location did this occur; where did the trip stop; what part was done by metro, bus or tram; how long did (s)he wait for a connection; did (s)he forgot to mention that (s)he carpools, or goes to buy groceries on the way home etc.). In light of this recognised potential, a growing body of work has been done in recent years. Particularly, the potential to use Global Navigation Satellite System (GNSS) data in replacing, or improving, travel surveys with crowdsourced data has been investigated (Vij and Shankari, 2015; Calabrese et al., 2013; Hasan and Ukkusuri, 2014). Furthermore, a more detailed look in travel time estimations indicated that penetration rates for GNSS-based probe data are now suitable for travel time estimation (Patire et al., 2015). Another promising data source for mobility studies is seen in mobile network data, particularly call detail records (CDR) (Toole et al., 2015) and positioning data (Chen et al., 2015). These are most often used for deriving trip's origin and destination locations (Alexander et al., 2015; Iqbal et al., 2014) and traffic zones extraction (Dong et al., 2015). And although the applicability of big data in the field of travel behaviour analysis has gained much attention, these insights are by far unexplored in current existing segmentation theories.

Extending the current research on crowdsourced mobility data and application of market segmentation techniques in the field of mobility, and in the meantime addressing the above mentioned limitations, the fundamental research contributions of this work can be situated in the following areas (i) in this paper we aim to bridge between psychological, social marketing and ICT research in field of mobility, by exploring how attitude based segments can be deduced from high resolution mobility behaviour data crowdsourced via mobile phones. For this purpose we will rely on the results of the applied theory of planned behaviour developed under European project SEGMENT (Anable and Wright, 2013). We do this as the attitude based segmentation has proven to have the highest potential on influencing people to adopt more sustainable means of transportation. In addition, results are well documented (Anable, 2005, 2013a,b; Ladbury, 2013) and as well as examples of ongoing and completed implementations for mobility management campaigns (Machado, 2015; Lassen Bue et al., 2013). In addition, (ii) we aim at bridging between applicability of small (survey based) and big data for mobility behaviour analysis. As recent literature well identifies this gap (Chen et al., 2015; Calabrese et al., 2013; Toole et al., 2015), big data are referred only as CDR and positioning data for mobility studies and potential of mobile sensed data (data collected based on dedicated smartphone apps) is being neglected. We investigate the potential to advance mobility research and bridge between small and big data based on mobile sensed data as they allow more seamless fusion between computer science research in this field (mainly focused on data analytics, but often without adequately recognising applicability context of achieved results for mobility needs) and transportation researchers who are well familiar with existing advances in the field of transportation but still new to advanced data analytics suitable for big data. And while CDR and positioning data exhibit major benefits as standardised forms or no additional cost for their collection (telecom operators collect these data anyway for billing purposes), for mobility studies they provide no contextual information and no ground truth information that could be used for development of learning algorithms. Mobile sensed data allow more balanced approach were traditional small data can be extended in the form of on-line surveys (Ji et al., 2015) and, in this way, big data supported with matched context. (iii) We extend current level of knowledge on attitude based segments by providing crowdsourced insights into their observed high resolution mobility behaviour, whereas so far studies have mainly tried to relate self-reported mobility behaviour with persons' attitudes towards different mobility options (Molin et al., 2016; Spears et al., 2013) and examine how variance in intention to use various travel modes can be explained based on these attitudes (Eriksson and Forward, 2011; Susilo et al., 2012). In addition, (iv) we explore potential to, after collection of test and learning datasets for target area, implement less resource demanding way to assign users in one of predefined attitudinal segments based on their high resolution observed mobility behaviour.

2. Method and data

2.1. 'Golden questions'

The SEGMENT project (Segment, 2013) aimed to test the use of consumer market segmentation techniques to help persuade people to change their travel behaviour and adopt more energy-efficient forms of transport. As mentioned in the introduction, this approach relies on surveys for data collection. Literature (Segment, 2013) reports that during the project more than 10,000 comprehensive attitudinal surveys have been completed by users in different cities containing more than 100 different questions. Based on these results, respondents were split in two groups (those with access to car and those without) and for each group, a hierarchical cluster procedure (Wards method, squared Euclidian distances) has been applied. For each cluster, the mean value for each variable has been used to start a *k*-means procedure (no-update method) in order to reduce the set to eight final clusters or consumer segments. In order to come to a practical approach, the size of the questionnaire still needed to be reduced. For this reason, the authors examined the segments based on the full list of survey questions and applied discriminant analysis to identify a smaller number of the most 'powerful' questions (those that discriminate the most among different segments of transportation system users). This smaller number of survey questions is referred to, by authors (Anable and Wright, 2013), as 'Golden questions' (Table 1). The accuracy of the 'Golden questions' (proportion of cases that were allocated to the correct segment) was examined using cross-validation and found to be higher than 82.5% and therefore above that expected by chance (Anable and Wright, 2013). Additional confirmation of applicability came from the successful use in mobility management campaigns in a number of, different sized, cities as Almada, Portugal (Machado, 2015), Gdynia, Poland (Lassen Bue et al., 2013), London, UK (Frost, 2015) and Utrecht, Nederland (Anable, 2013a,b). More details on the 'Golden questions' methodology can be found in the SEGMENT project's supporting documentation (Anable and Wright, 2013; Ladbury, 2013; Anable, 2005).

Overall, the 'Golden questions' contain eighteen question where the first one separates respondents based on their access to a car and the remainder are attitudinal questions (mainly in the form of statements where respondents are asked whether they agree or disagree on a 5-point Likert scale). Furthermore, these questions are grouped in three categories based on their conceptual meaning: questions solely for those that have access to a car, for those that do not and questions common for all users (Fig. 1). For the purpose of our research, participants were able to fill in the 'Golden questions' survey via the smartphone application used to track their mobility behaviour, as part of their registration process.

After completing the survey the participants are allocated into one of eight resulting segments (Table 2). These reflect their attitudes towards car use, cycling, electric vehicles or wider issues such as climate change and health.

In the following sections, we explore the possibility to automatically detect these attitude based segments based on high resolution mobility behaviour data without the need for users to answer the 'Golden questions'.

2.2. Collection of high resolution mobility behaviour data

In order to collect high resolution mobility behaviour data, an Android smartphone application (Routecoach, 2015) has been used and data is stored using the MOVE data platform (Ghent University, 2015). The application was freely available and had a registration process with the option to fill in the 'Golden questions' survey. During the registration process, every user was assigned a unique ID number used for mobility behaviour tracking. For every user's ID more than one device ID could be added (every time the user would use his/hers ID to complete the installation of the smartphone application for

Table 1
The 'Golden questions' (Anable and Wright, 2013).

'Golden questions'
Q1: Have you driven a car or van in the past 12 months?
Q2: For most journeys, I would rather use the car than any other form of transport
Q3: I like to drive just for the fun of it
Q4: I am not interested in reducing my car use
Q5: Driving gives me a way to express myself
Q6: How likely are you to drive in the next 12 months?
Q7: I am not the kind of person who rides a bicycle
Q8: I feel I should cycle more to keep fit
Q9: I find cycling stressful
Q10: Cycling can be the quickest way to travel around
Q11: I like travelling by bicycle
Q12: I am not the kind of person that likes to walk a lot
Q13: I feel I should walk more to keep fit
Q14: I like travelling by walking
Q15: I am not the kind of person to use the public transportation
Q16: In general, I would rather cycle than use the bus
Q17: I feel a moral obligation to reduce my emissions of greenhouse gases
Q18: People should be allowed to use their cars as much as they like

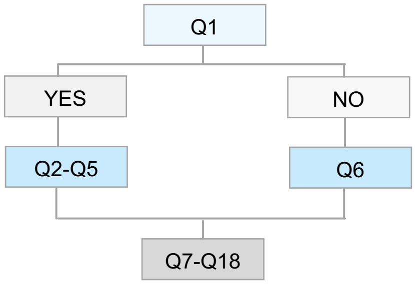

Fig. 1. Survey structure.

mobility tracking on a new device, a new ID would be automatically assigned to that device). After the installation and registration process, a user could either choose to register his mobility behaviour actively or allow data to be collected in background. Switching between the two modes was possible at any time. In the background data collection mode, the user's location and timestamp are continuously recorded with a frequency of 0.1 Hz. On the other hand, when a user chooses to actively register his mobility behaviour, the app basically functions as a mobile travel diary. By marking the transportation mode used for the trip, a user starts 'active' trip recording that collects location and time with a frequency of 1 Hz. The user also has an option to select one of the pre-offered trip purposes in the app but this is not obligatory. During the trip, the user is able to switch between transportation modes through the user interface, e.g. when walking to the train station and continuing the trip by the train, one would tap the 'pedestrian' icon at the beginning of the trip and swipe to 'train' at the train station. These parts of the trip travelled with different transportation modes we considered to be a trip segments and every of them had unique trip segment ID as well as every trip. Therefore, in the database, trip segment IDs that have the same trip ID are a part of one (and the same) trip and their order is recognised based on the timestamps. After reaching the destination, the user would tap the app to mark the end of the trip. From this data we extract the travel duration and distance. The complete list of variables collected is given in Table 3.

Overall, 1026 users installed the app and completed the registration process. In average, every 15th user had more than one device used for collection of mobility data (e.g. two smartphones or smartphone and tablet). Of our user base, we have used data from 629 users. These are users who have both completed the 'Golden questions' survey and have actively registered their mobility behaviour. Overall, 126,380 trips that satisfied these criteria have been collected over a period of six months (from Jan 2015 to Jun 2015) in the region of Flanders, Belgium.

In order to try to capture how attitude based segments can be deduced from the collected high resolution mobility behaviour data, we applied support vector machines. A short overview of this approach is given in the following section.

Table 2
Users segments (based on (Ladbury, 2013)).

Segment	Short description
Active aspirers	Have a high moral obligation to the environment and are highly motivated to use active transport modes, predominantly cycling as they believe that it is quick and provides freedom and fitness. They are not public transport users and see lots of problems with using it.
Carfree choosers	Do not like driving and think that cars lead to unhealthy lifestyles, they prefer cycling, public transport (do not think it is stressful or problematic) and walking. They feel a high moral obligation to the environment and are more likely to be women.
Car contemplators	They do not use car, have the highest proportion of non driving licence owners, but would like to as they see cars as status symbols. They see lots of problems with the public transportation use and find it, same as cycling, stressful. They believe walking is healthy and have a neutral or moderate attitude towards the environment.
Devoted drivers	Have no intention of reducing car use and think successful people use car. They do not use public transportation, nor cycling, and think walking is too slow. They are not motivated by fitness and have a very low moral obligation to the environment.
Image improvers	Like to drive, see the car as a way of self-expression and do not want to cut down car use. They do not use the public transportation but see cycling also as way of expressing them self's and a good way to keep fit. They have neutral or moderate environmental attitudes.
Malcontented motorists	They find driving stressful and have a moderately strong intention to reduce car use, but not to increase the use of public transport. Although, they would rather use the public transportation than cycle. They have a small level of environmental consciousness.
Practical travellers	They use car only when necessary as they think that it reduces the quality of life. They prefer cycling, as quicker, over the use of public transportation and would also walk when it seems more practical. They are not motivated by climate change and see local pollution and congestion as issues. They are highly educated and above-average part-time working.
Public transport dependents	They think people should be allowed to use cars and would like to see less congestion (they consider more roads as appropriate solution). They use public transport, although they think that it is not the quickest method. They do not cycle, but would like to walk more for fitness. They are not motivated by the environment, are least likely to start driving and have the highest number of retired people.

Table 3
Description of the crowdsourced mobility behaviour data.

Variable	Description
User ID	Unique user identification
Segment profile	Segment profile as defined by the 'Golden questions' survey
Trip ID	Unique trip identification
Trip segment ID	Unique trip segment identification
Device ID	Unique device identification
Trip start time	Trip starting time
Trip end time	Trip ending time
Trip duration	The difference between trip starting and ending time
Trip distance	Trip distance
Transportation mode	Transportation mode used
Trip purpose	Trip purpose
Trip starting point coordinates	Trip segment starting point coordinates
Trip ending point coordinates	Trip segment ending point coordinates
Indication	Indication whether the trip, purpose and transportation mode were collected in the background or actively registered (confirmed by the user)

2.3. Support vector machines

Support vector machines (SVM) belong to the supervised machine learning group of algorithms and can be used both for classification (Joo et al., 2015; Cavar et al., 2011) or regression analysis (Vlahogianni, 2015; Yu et al., 2011; Wang and Shi, 2013). For our purpose, automatically detecting attitude based segments using crowdsourced high resolution mobility behaviour data, we focus in this paper only on the SVM classification analysis as the attitudinal segments correspond to categorical dependent variables.

In general, the SVM classification is based on the concept of decision hyperplanes that define decision boundaries (separates between a set of objects having different class memberships e.g. belonging to the different market segments). Often this task is not simple. The use of structures more complex than linear ones is needed to correctly classify new objects (test data) on the basis of the examples that are available (training data). For this purpose different mathematical functions (kernels) can be used in order to map objects in the n dimensional space (Schlkopf and Smola, 2001; Hamel, 2011; James et al., 2013). In our example, when mapping categorical variables, we adopt dummy variables created with case values as either 0 or 1. Thus, a categorical dependent variable consisting of four levels, say (A, B, C, D), is represented by a set of four dummy variables:

$$A:\ \{1000\},\ B:\ \{0100\},\ C:\ \{0010\},\ D:\ \{0001\} \tag{1}$$

In addition, we divided the complete data set in two parts; 75% has been used as training and 25% as test dataset, with mutually exclusive user ID's. The input dataset consists of trip segment logs with actively registered user input. A data vector includes information on the user, trip, device and trip segment ID's, trip segment duration, distance and transportation mode as well as the start and end time and location. In addition, the attitudinal segments, based on the 'Golden questions' survey, are assigned to each user and used as the target (output) categorical variable (Table 4).

For the SVM classification we applied C-SVM type and one-against-all approach to map multiclass problem into binary classification problem. The reason for applying C-SVM type is the fact that crowdsourcing can result in large datasets, and we wanted to obtain a scalable runtime in regard to the number of input samples. The literature suggests that in this case C-SVM is a better option over, for example, nu-SVM classification (Chang and Lin, 2001). For the applied C-SVM type the minimization error function is defined as:

$$\frac{1}{2}w^T w + C\sum_{i=1}^{N}\xi_i \tag{2}$$

Subject to the constraints:

$$y_i\left(w^T\phi(x_i) + b\right) \geqslant 1 - \xi_i \tag{3}$$

$$\xi_i \geqslant 0 \tag{4}$$

where $i = 1, \ldots, N$, w is the vector of coefficients, C is the capacity constant, b is a constant, and ξ_i represents parameters for handling non-separable data (inputs). The index i labels the N training cases ($y \in \pm 1$ represents the class labels and x_i represents the independent variables). The ϕ stands for kernel function (radial basis function - RBF) that transforms input to the feature space:

$$K(X_i, X_j) = \phi(X_i) \cdot \phi(X_j) = \exp\left(-\gamma|X_i - X_j|^2\right) \tag{5}$$

Table 4
Model variables.

Variable	Variable type	Type	Length	Example value	Units or categorical values
Used ID	Independent	Double	8	168	No unit
Trip ID	Independent	Text	36	26cfc8e7-7d40	No unit
Trip segment ID	Independent	Text	36	439d-8c48	No unit
Device ID	Independent	Integer	8	1085	No unit
Trip segment start time	Independent	Double	8	15:15:55	hh:mm:ss
Trip segment end time	Independent	Double	8	15:25:51	hh:mm:ss
Trip segment duration	Independent	Integer	4	9	minutes
Trip segment distance	Independent	Integer	6	1016	meters
Transportation mode	Independent	Text	13	BIKE	Bike; Car[a]; Foot; Public transportation (bus); Train
Trip purpose	Independent	Text	10	WORK	Drop-off; Home; Other; Recreation; Shop; Visit; Work; null
Trip segment starting point X coordinate	Independent	Double	8	4.1234	degrees (WGS84)
Trip segment starting point Y coordinate	Independent	Double	8	50.1234	degrees (WGS84)
Trip segment ending point X coordinate	Independent	Double	8	4.1234	degrees (WGS84)
Trip segment ending point Y coordinate	Independent	Double	8	50.1234	degrees (WGS84)
Segment profile	Dependent	Text	27	Practical travellers	Active aspirers; Carfree choosers; Car Contemplators; Devoted drivers; Image improvers; Malcontented motorists; Practical travellers; Public transport dependents

[a] Both Car (as driver) and Car (as passenger) are labelled as Car.

It is not possible to known beforehand which capacity constants C (Eq. (1)) and γ (Eq. (4)) are best for a given problem. Because their value is important to keep the training error small and in order to generalize well (Anguita and Oneto, 2011), we applied the incremental grid-search on C (with step 1; range from 1 to 10) and γ (with step 1/30; range from 1/30 to 1/2) and those that had the best average 10-fold cross-validation accuracy are the ones chosen for use on the test data. The obtained value for C was 5 and for γ 1/6.

For the v-fold cross-validation, the total number of cases are divided into v (in our example $v = 10$) sub samples $Z_1, Z_2, \ldots, Z_v$ of almost equal sizes $N_1, N_2, \ldots, N_v$, respectively. The v-fold cross-validation estimate is the proportion of cases in the subsample Z that are misclassified by the classifier constructed from the subsample $Z - Z_v$. This estimate is computed in the following way:

$$R(d^{(v)}) = \frac{1}{N_v} \sum_{(x_n, j_n) \in Z_v} X(d^{(v)}(x_n) \neq j_n) \tag{6}$$

where $d^{(v)}(x)$ is the classifier computed from the sub sample $Z - Z_v$ and X is the indicator function for which it is valid:

- $X = 1$, if the statement $X(d^{(v)}) \neq j_n)$ is true.
- $X = 0$, if the statement $X(d^{(v)}) \neq j_n)$ is false.

For the test sample estimate, the accuracy is calculated as follows: as the total number of cases was divided into two subsamples Z_1 (training dataset) and Z_2 (test dataset - not used for constructing the classifier). The test sample estimate is the proportion of cases in the test dataset which are misclassified by the classifier constructed from the learning dataset. This estimate is computed in the following way:

$$R(d) = \frac{1}{N_2} \sum_{(x_n, j_n) \in Z_2} X(d(x_n) \neq j_n) \tag{7}$$

Since the input vector was trip based, and not user based, for every user the attitudinal segment estimation is deducted based on the majority vote for all of the trips (s)he made.

3. Results

3.1. Observed mobility behaviour

As previously mentioned in Section 2.2, 126,380 trips have been collected over Flanders. Fig. 2 shows trips made by a car or bike, highlighting the most used routes and areas.

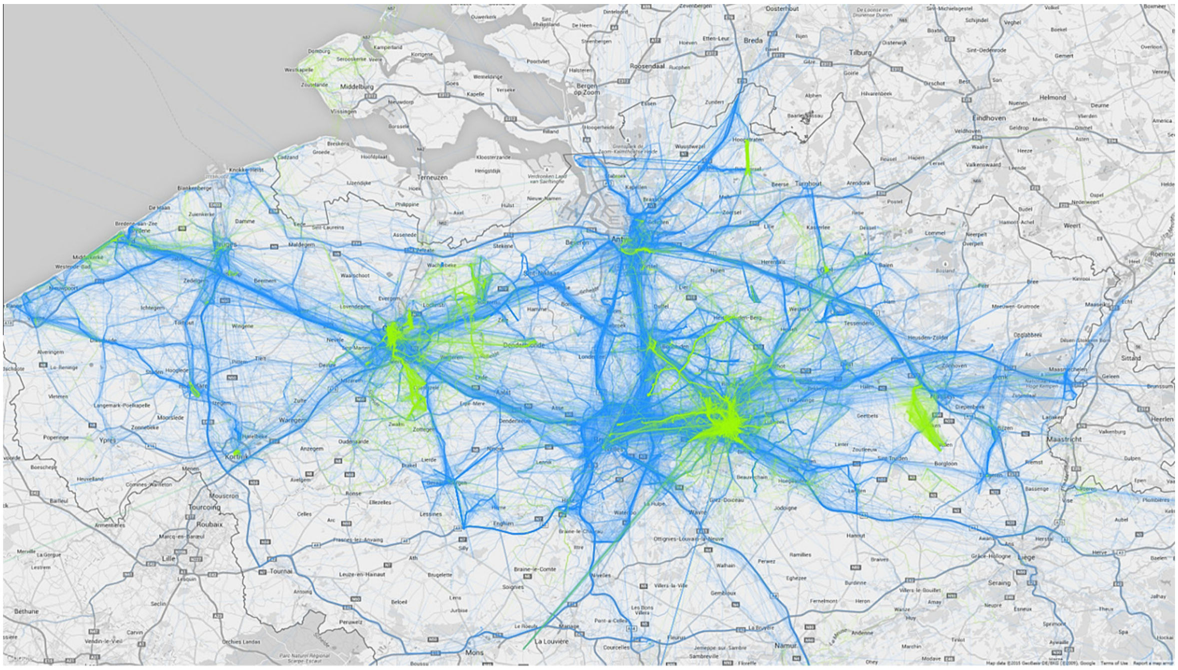

Fig. 2. Sample of crowdsourced data (blue – car trips, green – bike trips). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

In our data sample six different transportation modes have been reported by users:

- Bike
- Car (as driver)
- Car (as passenger)
- Foot
- Public transportation (bus)
- Public transportation (train).

Among these, bike was used the most and public transportation and car (as passenger) the least (Fig. 3). Considering the observed trip purposes, in some cases the trip purpose has not been identified by the user. In this case a 'null' value (Fig. 4) is reported. For the trip purposes, most of the trips were related to home or going to work activities.

Considering trip distances (Fig. 5), the longest trip was up to 217 km, but most of the trips were less than 5 km (64%) while 20% of them were shorter than 1000 m. The distribution of trips during the day (Fig. 6) clearly indicate the peak hours in the morning, between 7 and 9 h, and an afternoon peak between 16 and 19 h. During these time periods almost half of the recorded trips were made.

3.2. Mapping of the attitude based mobility segments on high resolution mobility behaviour data

When applying the SVM to detect attitude based segments using crowdsourced high resolution mobility behaviour data, the model was able to successfully identify to which attitudinal segment a user belonged in 98.9% of the cases. The cross-validation accuracy was 97%, while success rate for learning sample was 99.6% and for the test sample 96.7%. The average error (percentage of misclassified users) per attitudinal segment was 1.5% and its standard deviation 2.2 (Table 5). The confusion only happened for seven users (Fig. 7) - between Active aspirers, Practical travellers, Image improvers and Malcontent motorist segments (Table 6).

Taking into the account relative observations, the confusion was the highest for the Image improvers (5.88%), misclassifying them as Practical travellers. This was the only users' segment where the success rate was less than 95%. On the other hand, for Car-free choosers, Devoted drivers, Public transport dependents and Car contemplators the success rate was the highest.

Considering the complexity of the n dimensional space, overall 286 support vectors were needed to create boundaries between the mapped behaviour observations belonging to different attitude based segments. The most challenging part

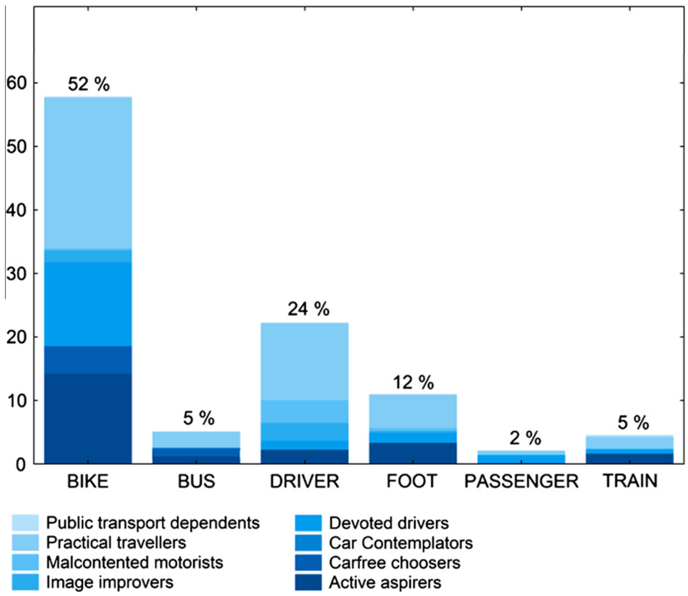

Fig. 3. Usage of transportation modes.

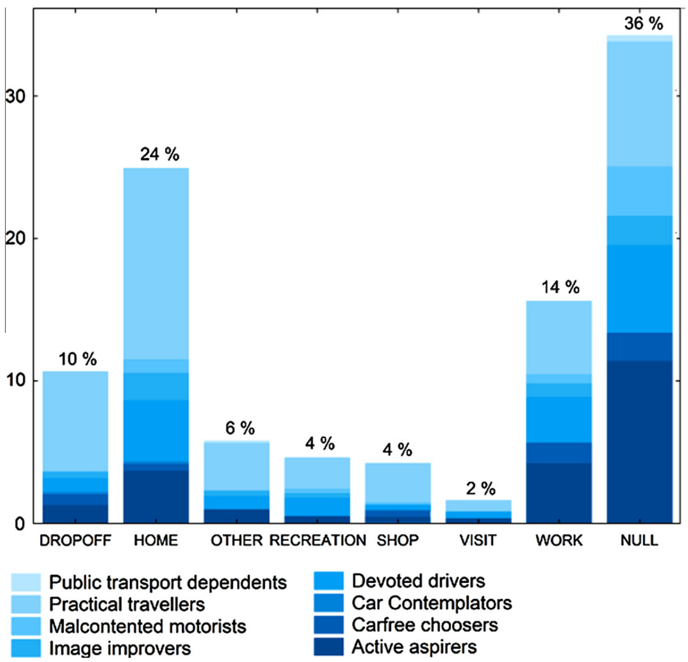

Fig. 4. Trip purpose.

was to separate Car-free choosers segment as this required 107 support vectors and the least challenging was to identify the boundaries of n dimensional space where Active aspirers observations were mapped (Table 7).

4. Discussion

We have found that attitude based segments can be successfully captured using high resolution mobility behaviour data crowdsourced via active smartphone devices. The applied SVM classification algorithm yields a success rate of 98.9% in predicting the results of the 'Golden questions' survey. The 'Golden questions' are smaller set of questions, extracted from the

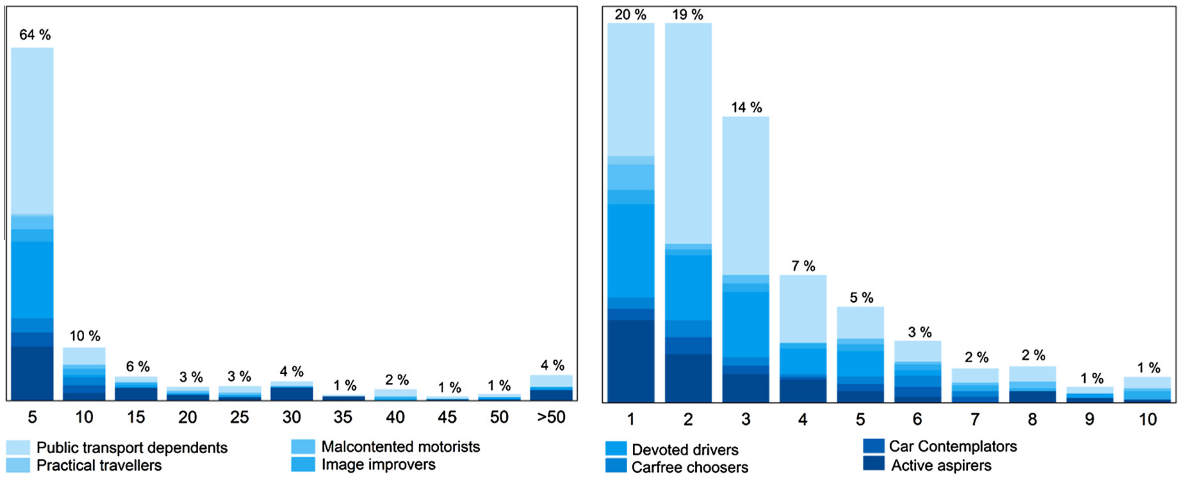

Fig. 5. Distribution of trips distances in km (left) and distribution of trip distances shorter than 10 km (right).

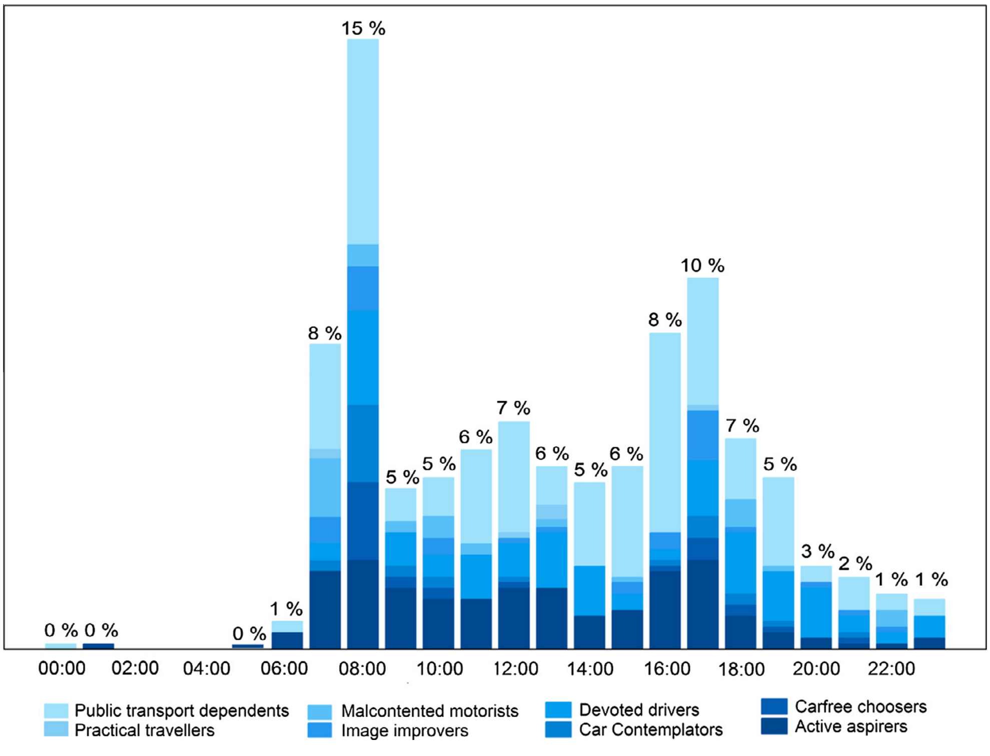

Fig. 6. Distribution of time of day when trip was stated.

list of more than 100 questions used to define eight attitudinal segments, that discriminate the most among different segments of transportation system users. As a reduced set of questions, 'Golden questions' themselves are not able to fully explain all attitudes that differentiate between eight segments but are developed for practical reasons and confirmed to have satisfying level of accuracy (as explained in Section 2). Furthermore, their applicability is already confirmed by the successful implementation in number of existing mobility management campaigns. Thus we would interpret our model as having high accuracy in predicting the results of 'Golden questions' survey and applicable for practical use as an automated replacement for attitudinal segmentation which would otherwise be based on the 'Golden questions' paper surveys. As mentioned in the introduction, the attitude based segmentation is considered to be more sophisticated than a solely behaviour based one. Nevertheless, although we rely on behaviour data, our findings do not take a step back. By using high resolution data, we take a step forward in mapping the attitude based segments by applying the advanced and innovative ways to collect data

Table 5
Classification summary.

	Total	Correct	Incorrect	Correct (%)	Incorrect (%)
Active aspirers	138	136	2	98.55	1.45
Carfree choosers	28	28	0	100.00	0.00
Devoted drivers	114	114	0	100.00	0.00
Image improvers	34	32	2	94.12	5.88
Malcontented motorists	28	27	1	96.43	3.57
Practical travellers	281	279	2	99.29	0.71
Public transport dependents	3	3	0	100.00	0.00
Car contemplators	3	3	0	100.00	0.00

on mobility behaviour via smartphone devices. A far more detailed set of information regarding individuals' mobility behaviour can be captured compared to the standard but widely used survey techniques for mobility oriented segmentation. Our findings go in line with those of Prillwitz and Barr (2011), who, based on the survey input of about 1500 individuals, applied two segmentation approaches aiming to identify gaps between different domains of individual travel behaviour and varying role of attitudes for travel decisions. Among their findings, they concluded that attitudinal segments did not differ substantially in terms of socio-demographic and socio-economic characteristics, but possessed very different behavioural daily travel patterns. Our research, based just on these travel patterns (although collected differently and with far higher resolution than travel patterns captured via surveys), successfully identified support vectors that separate attitudinal segments with very high precision. In addition, as one of a kind research we aimed at bridging between (i) psychological studies attempting to identify the typical characteristics of those people who are interested in travel behaviour change (Hunecke et al., 2010; Anable, 2005; Prillwitz and Barr, 2011), (ii) transportation policy studies attempting to identify which measures are more likely to gain desired impact in the sustainable mobility sense (Bamberg et al., 2011; Lyons et al., 2008; Cairns et al., 2008), (iii) social marketing studies trying to identify what message is most likely to trigger desired travel behaviour change (McKenzie-Mohr, 2000; Kassirer and Lagarde, 2010) and (iv) smart city studies that try to take the advantage of innovative information and communication (ICT) tools in ensuring data-driven smart mobility management (Motta et al., 2013;

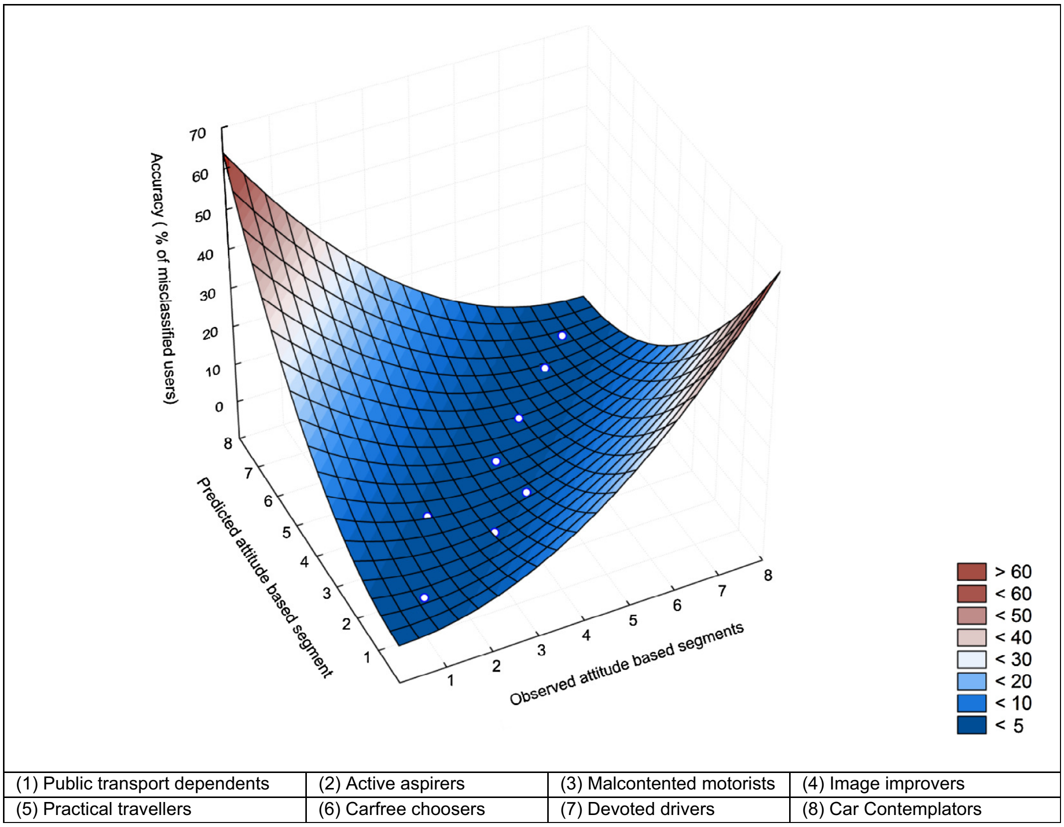

Fig. 7. Prediction results.

Table 6
Confusion matrix.

	Active aspirers	Carfree choosers	Devoted drivers	Image improvers	Malcontent motorists	Practical travellers	Public transport dependents	Car contemplators
Active aspirers	136	0	0	0	0	2	0	0
Carfree choosers	0	28	0	0	0	0	0	0
Devoted drivers	0	0	114	0	0	0	0	0
Image improvers	0	0	0	32	0	2	0	0
Malcontented motorists	0	0	0	0	27	1	0	0
Practical travellers	2	0	0	0	0	279	0	0
Public transport dependents	0	0	0	0	0	0	3	0
Car contemplators	0	0	0	0	0	0	0	3

Table 7
Model summary.

Number of support vectors	286
Number of support vectors (Active aspirers)	3
Number of support vectors (Car-free choosers)	107
Number of support vectors (Devoted drivers)	23
Number of support vectors (Image improvers)	67
Number of support vectors (Malcontented motorists)	14
Number of support vectors (Practical travellers)	8
Number of support vectors (Public transport dependents)	41
Number of support vectors (Car contemplators)	23

Semanjski and Gautama, 2015). Additionally, our research presents one of a kind in the field of mobility data collection, particularly through exploring the potential of ICT versus traditional surveys and bridging between small and big data use for mobility behaviour studies. Whereas so far the focus was mainly on analysing differences between survey-reported and global positioning system (GPS) recorded trips (Bricka et al., 2012; Murakami and Wagner, 1999; Chen et al., 2010), we try to explore the crowdsourcing potential and its role in facilitating data-driven mobility management. Particularly, we consider that our findings close the gap between the SEGMENT project findings (Anable and Wright, 2013), that identified eight attitudinal segments for mobility campaigns, and intentions of on-going SUPERHUB project (SUPERHUB, 2015) to compose personalised notification messages, for these eight segments, to be delivered via smartphones with the aim to trigger/facilitate travel behaviour changes towards the more sustainable one (Forbes et al., 2014). We do this by providing a tool to match attitudinal segments with smartphone users, based solely on the crowdsourced data and without any additional effort by users. This way, seamlessly, personalised mobility campaigns can reach wider population, on a less resource demanding way (Bohte and Maat, 2009), with greater effect as, potentially, entire population can be mapped into respective attitudinal segments and not just the survey participants. This area particularly seems interesting in the context of smart mobility management and future case studies to confirm this concept are desired. In addition, since approach is based on the crowdsourced data with time-space indication, targeted messages could be potentially delivered at locations where users, belonging to specific segment, are located at specific time interval (e.g. variable message signs during peak hour at city centre, parking entry or public transportation locations).

5. Conclusion

In this paper we successfully mapped sustainable mobility oriented attitudinal based segments to the *n* dimensional space defined by crowdsourced mobility behaviour variables. This way we have confirmed previous indications that different attitudinal segments show different daily travel patterns. In addition, we have bridged different research fields in order to facilitate the practical implementation of psychological, social marketing and ICT findings in the field of smart city mobility management. We have proposed a new means of mobility data collection to identify groups of people that are likely to be more responsive to different policy options such as pricing measures, traffic calming or advertising campaigns. The method is shown to be accurate as well as less time consuming to the user and more resource efficient for the transportation research.

Acknowledgments

The authors are indebted to Rik Bellens and Angel Lopez Aguirre (Ghent University) for technical support and database management, as well as to Elke Franchois and Vicky Dierckx (Mobiel21) for communication campaign and engagement of research participants.

This research is funded by INTERREG North-West Europe project New Integrated Smart Transport Options (NISTO), the Flemish government agency for Innovation by Science and Technology and the Flemish Institute for Mobility.

References

Ajzen, I., 1991. The theory of planned behavior. Organ. Behav. Hum. Dec. 50 (2), 179–211.

Alexander, L., Jiang, S., Murga, M., González, M.C., 2015. Origin–destination trips by purpose and time of day inferred from mobile phone data. Transport. Res. Part C – Emer. 58, 240–250.

Anable, J., 2005. 'Complacent Car Addicts' or 'Aspiring Environmentalists'? Identifying travel behaviour segments using attitude theory. Transp. Pol. 12 (1), 65–78.

Anable, J., 2013a. SEGMENT Survey No 2: Data Evaluation Report – Almada. University of Aberdeen, Aberdeen.

Anable, J., 2013b. SEGMENT Survey No 2: Data Evaluation Report – Utrecht. University of Aberdeen, Aberdeen.

Anable, J., Wright, S., 2013. Segment. <http://www.segmentproject.eu/hounslow/segment.nsf/Files/SFF-266/$file/Deliverable%207.8.4%20GOLDEN%20QUESTIONS%20AND%20SOCIAL%20MARKETING%20GUIDANCE%20REPORT.pdf> (accessed 14 March 2014).

Anguita, D., Oneto, L., 2011. In-Sample Model Selection for Support Vector Machines. CA, IEEE, San Jose.

Bamberg, S., Fujii, S., Friman, M., Garling, T., 2011. Behaviour theory and soft transport policy measures. Transp. Pol. 18, 228–235.

Bamberg, S., Hunecke, M., Blöbaum, A., 2007. Social context, personal norms and the use of public transportation: two field studies. J. Environ. Psychol. 27 (3), 190–203.

Bohte, W., Maat, K., 2009. Deriving and validating trip purposes and travel modes for multi-day GPS-based travel surveys: a large-scale application in the Netherlands. Transport. Res. Part C – Emer. 17 (3), 285–297.

Bricka, S.G., Sen, S., Paleti, R., Bhat, C.R., 2012. An analysis of the factors influencing differences in survey-reported and GPS-recorded trips. Transport. Res. Part C – Emer. 21, 67–88.

Cairns, S. et al., 2008. Smarter choices: assessing the potential to achieve traffic reduction using 'soft measures'. Transport. Rev. 28 (5), 593–618.

Calabrese, F. et al., 2013. Understanding individual mobility patterns from urban sensing data: a mobile phone trace example. Transport. Res. Part C – Emer. 26, 301–313.

Cavar, I., Kavran, Z., Petrovic, M., 2011. Hybrid approach for urban roads classification based on GPS tracks and road subsegments data. PROMET 23 (4), 289–296.

Chang, C.-C., Lin, C.-J., 2001. Training ν-support vector classifiers: theory and algorithms. Neural Comput. 13, 2119–2147.

Chen, C., Gong, H., Lawson, C., Bialostozky, E., 2010. Evaluating the feasibility of a passive travel survey collection in a complex urban environment: lessons learned from the New York City case study. Transport. Res. Part A – Pol. 44 (10), 830–840.

Chen, C. et al., 2015. The promises of big data and small data for travel behavior (aka human mobility) analysis. Transport. Res. Part C – Emer. 68, 285–299.

Cheng, H., Luo, H.-Y., Zhao, F., 2012. Device-clustering algorithm in crowdsourcing-based localization. J. China Univ. Posts Telecommun. 19 (2), 114–121.

Delang, C.O., Cheng, W.-T., 2012. Consumers' attitudes towards electric cars: a case study of Hong Kong. Transport. Res. Part D – Tre. 17 (6), 492–494.

Diana, M., Pronello, C., 2010. Traveler segmentation strategy with nominal variables through correspondence analysis. Transp. Pol. 17 (3), 183–190.

Dong, H. et al., 2015. Traffic zone division based on big data from mobile phone base stations. Transport. Res. Part C – Emer. 58, 278–291.

Eriksson, L., Forward, S.E., 2011. Is the intention to travel in a pro-environmental manner and the intention to use the car determined by different factors? Transport. Res. Part D – Tre. 16, 372–376.

Forbes, P. et al., 2014. Towards using segmentation-based techniques to personalize mobility. EAI Endorsed Trans. Ambient Syst. 4, e4.

Frost, M., 2015. The Urban Mobility Observatory. <http://www.eltis.org/discover/case-studies/segment-employee-campaign-london-hounslow-uk> (accessed 22 June 2015).

De Jong, G., Gunn, H., Ben-Akiva, M., 2004. A meta-model for passenger and freight transport in Europe. Transp. Pol. 11 (4), 329–344.

Ghent University, 2015. MOVE. <http://move2.ugent.be> (accessed 05 May 2015).

Gong, H., Chen, C., Bialostozky, E., Lawson, C.T., 2012. A GPS/GIS method for travel mode detection in New York City. Comput. Environ. Urban 36 (2), 131–139.

Goodchild, M.F., Li, L., 2012. Assuring the quality of volunteered geographic information. Spatial Stat. 1, 110–120.

Hamel, L.H., 2011. Knowledge Discovery with Support Vector Machines (Wiley Series on Methods and Applications in Data Mining), first ed. Wiley-Interscience, Hoboken.

Hasan, S., Ukkusuri, S.V., 2014. Urban activity pattern classification using topic models from online geo-location data. Transport. Res. Part C – Emer. 44, 363–381.

Hunecke, M., Haustein, S., Böhler, S., Grischkat, S., 2010. Attitude-based target groups to reduce the ecological impact of daily mobility behavior. Environ. Behav. 42 (1), 3–43.

Hu, X., Wang, J., Wang, L., 2013. Understanding the travel behavior of elderly people in the developing country: a case study of Changchun, China. Proc. – Soc. Behav. Sci. 96 (6), 873–880.

IEA, 2014. Key World Energy Statistics. International Energy Agency, Paris.

Iqbal, S., Choudhury, C.F., Wang, P., González, M.C., 2014. Development of origin–destination matrices using mobile phone call data. Transport. Res. Part C – Emer. 40, 63–74.

James, G., Witten, D., Hastie, T., Tibshirani, R., 2013. An Introduction to Statistical Learning: With Applications in R (Springer Texts in Statistics), first ed. Springer, Berlin.

Ji, Y., Mishalani, R.G., McCord, M.R., 2015. Transit passenger origin–destination flow estimation: efficiently combining onboard survey and large automatic passenger count datasets. Transport. Res. Part C – Emer. 58, 178–192.

Joo, S., Oh, C., Jeong, E., Lee, G., 2015. Categorizing bicycling environments using GPS-based public bicycle speed data. Transport. Res. Part C – Emer. 56, 239–250.

Kassirer, J., Lagarde, F., 2010. Changing Transportation Behaviour. Transport Canada, Ottawa.

Ladbury, P., 2013. The Segment Toolkit. <http://www.segmentproject.eu/hounslow/segment.nsf/Files/SFF-318/$file/Deliverable%207-8.3%20Social%20Marketing%20Toolkit.pdf> (accessed 19 February 2014).

Lassen Bue, E., Makowski, W., Reiter, K., 2013. BYPAD Certification Report - City of Gdynia. Grupa Inspro, Gdynia.

Lyons, G. et al., 2008. Public Attitudes to Transport: Knowledge Review of Existing Evidence. Centre for Transport and Society, University of West England, Bristol.

Machado, P., 2015. New Residents Mobility Welcome Kit. Almada. Portugal. <http://www.eltis.org/discover/case-studies/new-residents-mobility-welcome-kit-almada-portugal> (accessed 22 May 2015).

McKenzie-Mohr, D., 2000. Promoting sustainable behavior: an introduction to community-based social marketing. J. Soc. Iss. 56 (3), 543–554.

Molin, E., Mokhtarian, P., Kroesen, M., 2016. Multimodal travel groups and attitudes: a latent class cluster analysis of Dutch travelers. Transport. Res. Part A – Pol. 83, 14–29.

Motta, G. et al., 2013. Integrated mobility: a research in progress. J. Softw. Eng. Appl. 6, 97–101.

Murakami, E., Wagner, D., 1999. Can using global positioning system (GPS) improve trip reporting? Transport. Res. Part C – Emer. 7 (2–3), 149–165.

Patire, A.D., Wright, M., Prodhomme, B., Bayen, A.M., 2015. How much GPS data do we need? Transport. Res. Part C – Emer. 58, 325–342.

Prillwitz, J., Barr, S., 2011. Moving towards sustainability? Mobility styles, attitudes and individual travel behaviour. J. Transp. Geogr. 19 (6), 1590–1600.

Redmond, L.S., 2000. Identifying and Analyzing Travel-Related Attitudinal, Personality, and Lifestyle Clusters in the San Francisco Bay Area. Institute of Transportation Studies, University of California, Davis.

Routecoach, 2015. Routecoach. <http://www.routecoach.be/> (accessed 20 June 2015).

Schlkopf, B., Smola, A.J., 2001. Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond (Adaptive Computation and Machine Learning), first ed. The MIT Press, Cambridge.

Segment, 2013. Segmented Marketing for Energy Efficient Transport. <http://www.segmentproject.eu/> (accessed 22 February 2015).

Semanjski, I., Gautama, S., 2015. Smart city mobility application – gradient boosting trees for mobility prediction and analysis based on a crowdsourced data. Sensors-Basel 15 (7), 15974–15987.

Spears, S., Houston, D., Boarnet, M.G., 2013. Illuminating the unseen in transit use: a framework for examining the effect of attitudes and perceptions on travel behavior. Transport. Res. Part A – Pol. 58, 40–53.

SUPERHUB, 2015. SUstainable and PERsuasive Human Users moBility in Future Cities. <http://superhub-project.eu/> (accessed 01 June 2015).

Susilo, Y.O., Williams, K., Lindsay, M., Dair, C., 2012. The influence of individuals' environmental attitudes and urban design features on their travel patterns in sustainable neighborhoods in the UK. Transport. Res. Part D – Tre. 17, 190–200.

Toole, J.L. et al., 2015. The path most traveled: travel demand estimation using big data resources. Transport. Res. Part C – Emer. 58, 162–177.

Vij, A., Shankari, K., 2015. When is big data big enough? Implications of using GPS-based surveys for travel demand analysis. Transport. Res. Part C – Emer. 56, 446–462.

Vlahogianni, E.I., 2015. Optimization of traffic forecasting: Intelligent surrogate modeling. Transport. Res. Part C – Emer. 55, 14–23.

Wang, J., Shi, Q., 2013. Short-term traffic speed forecasting hybrid model based on Chaos-Wavelet Analysis-Support Vector Machine theory. Transport. Res. Part C – Emer. 27, 219–232.

Wu, X., Liu, H.X., 2014. Using high-resolution event-based data for traffic modeling and control: an overview. Transport. Res. Part C – Emer. 42, 28–43.

Yu, B., Lam, W.H., Lam Tam, M., 2011. Bus arrival time prediction at bus stop with multiple routes. Transport. Res. Part C – Emer. 19 (6), 1157–1170.