Developing an optimised activity type annotation method based on classification accuracy and entropy indices

Wim Ectors[a], Sofie Reumers[a], Won Do Lee[b,c*], Keechoo Choi[b], Bruno Kochan[a], Davy Janssens[a], Tom Bellemans[a] and Geert Wets[a]

[a] *Hasselt University, Transportation Research Institute (IMOB), Agoralaan, 3590 Diepenbeek, Belgium;*

[b] *Department of Transportation Engineering, Ajou University, 16499 Suwon, Republic of Korea*

[c] *Crime and Well-being Big Data Centre, Manchester Metropolitan University, M15 6BH Manchester, England*

* Corresponding author: Won Do Lee (Tel: +44-161-247-6538, Email: w.lee@mmu.ac.uk)

Developing an optimised activity type annotation method based on classification accuracy and entropy indices

The generation of substantial amounts of travel and mobility related data has spawned the emergence of the era of big data. However, this data generally lacks activity-travel information such as trip purpose. This deficiency led to the development of trip purpose inference (activity type imputation / annotation) techniques, of which the performance depends on the available input data and the (number of) activity type classes to infer. Aggregating activity types strongly increases the inference accuracy and is usually left to the discretion of the researcher. As this is open for interpretation, it undermines the reported inference accuracy.

This study developed an optimised classification methodology by identifying classes of activity types with an optimal balance between improving model accuracy, and preserving activity information from the original data set. A sensitivity analysis was performed. Additionally, several machine learning algorithms are experimented with. The proposed method may be applied to any study area.

Keywords: activity classification; activity class optimisation; big data annotation; trip purpose imputation; classification algorithms; activity entropy

Introduction

The contemporary collection of substantial amounts of travel and mobility related data has spawned the emergence of the era of big data. Mobility data is currently generated in real-time. The continuous advancements in data collection technology, reduced data storage costs and novel data processing procedures have enabled this evolution.

In the scientific realm, big transport data (more often than traditional transport data) are frequently used to estimate individual mobility patterns using *spatiotemporal* characteristics prominent in the data, in order to better understand human travel behaviour and transport users' needs. Furthermore, these estimations can provide pertinent insights to policy makers and transport service providers.

Despite the well-described strengths of big transport data in recent literature, some limitations specific to big transport data are encompassed as well. Big transport data recurrently lack activity-travel information, such as trip purpose, mode, and socio-demographic information. This deficiency has inspired researchers in the field to progress towards the development of annotation (inference) techniques, i.e. deducing people's activity-travel information from the observed, small-scale (survey) data and adopting the obtained rules to the big data sets.

The current study employs data mining methods in order to estimate activity types (sometimes otherwise denoted as trip or travel purposes, activity classes, activity categories or activity encoding)[a]. For this, temporal attributes of the observed (reference)

[a] Various terms are used in different research streams, but are in essence the same. Behavioural mining research may prefer to use 'activity type classes', whilst 'activity categories' could be used more in travel demand forecasting research. 'Activity encoding' may be used in any computational research.

data are employed. Data mining methods were applied in two steps, in which the first step is the search for an optimal number and composition of activity type classes (i.e. the formation of groups of activity types). In the second step, the activity type classes are modelled using probability and rule-based heuristic techniques. A sensitivity analysis was conducted on this optimisation strategy. The behavioural data mining process results can be utilised to infer the activity type classes in (big) transport data. This is a topic for future research.

In the next chapter, an overview of relevant literature is provided. The literature review covers three related topics, i.e. the opportunities of big transport data, behavioural data mining methods for activity categorisation and classification, and the challenges associated with these data mining methods. In the subsequent chapter, the different steps of the research process are described. Next, the experimental results are presented and a sensitivity analysis is provided. A conclusion finalises this paper.

Literature review

The opportunities of big transport data

The era of big data commenced with the appearance of large amounts of data, which are rich in e.g. spatiotemporal information and may be collected in real-time. Kitchin (2013) denotes this phenomenon as a 'data deluge'. The collection of daily trajectory data has been improved by advancements in ICT and by the improvement of location-aware technologies.

New transport data collection methods facilitate a better investigation and understanding of human travel behaviour, and provide much more sophisticated, wider-scale, finer-grained, real-time data. In recent big transport data researches, advanced technologies allow to exploit the collected long-term movement data, e.g. the use of mobile call detail records (CDRs). CDRs are not only employed to explore the details of individual mobility variability (Järv, Ahas, and Witlox 2014; Toole et al. 2015; Wang, He, and Leung 2017), but also to record individual 24h activity-travel sequences (Liu et al. 2014). Several CDR studies focused on the estimation of OD matrices (e.g. Alexander et al., 2015). Smartphones are another promising technology with respect to new travel survey methods. Zhao et al. (2015) present an exploratory analysis of the combination of a smartphone-based travel survey and the traditional household interview travel survey (HITS) in Singapore during 2012 and 2013. An interesting observation from this study is the lower amount of under-reporting of trips and the better accuracy in smartphone-based travel surveys compared to traditional surveys. Moreover, the increasing penetration rate of smartphones will decrease data collection costs and the absence of user burden will facilitate the collection of longitudinal data.

Despite the strengths of big transport data, major concerns regarding privacy (Graham and Shelton 2013; Kitchin 2013) and sampling bias (Yue et al. 2014) remain. In addition, analysing or mining big data implies large costs in terms of computation time and hardware resources.

Big transport data mining methods

As mentioned above, big transport data is very effective in exploring individual mobility patterns, i.a. due to the availability of temporal information (e.g. time stamps). However, the lack of adequate information about personal and activity-travel characteristics is a disadvantage (W. Do Lee et al. 2015). To overcome this limitation, recent studies develop novel behavioural data mining methods to deduce activity-travel information from surveys or from advanced location-containing data sets having spatial and temporal information (e.g. activity start times, activity durations, land use information) (Gong et al. 2014; Wolf, Guensler, and Bachman 2001; Lu, Zhu, and Zhang 2012; S. Lee and Hickman 2014; Shen and Stopher 2013; Montini et al. 2014; Lu and Zhang 2015; Simas-Oliveira et al. 2014; Lu, Zhu, and Zhang 2013; Feng and Timmermans 2015; Nurul Habib and Miller 2009). In the field of ITS (Intelligent Transportation Systems), transportation data mining methods that aim to integrate different data sources are generally denoted as data fusion (DF) techniques. Three different approaches are applied: statistical, probabilistic, and artificial intelligence approaches (Faouzi, Leung, and Kurian 2011).

Indeed, many studies are devoted towards the development of optimal data mining methods, especially estimating the activity types based on their behavioural attributes. The next section deals with the activity class categorisation and provides a review of activity classification methods that are available in the literature.

Activity class categorisation

As a proof of concept, Wolf et al. (2001) inferred trip purposes from GPS and land use related information. Many researches followed and advanced this domain of activity type annotation. Some focused on finding more and better data (Lu, Zhu, and Zhang 2012; Shen and Stopher 2013; Montini et al. 2014; Lu and Zhang 2015; Simas-Oliveira et al. 2014; Gong et al. 2014), others experimented with different methodologies and data mining algorithms (S. Lee and Hickman 2014; Simas-Oliveira et al. 2014; Lu, Zhu, and Zhang 2013). The activity type classes which are predicted (and the size of this set of classes) strongly impact the classification accuracy. Some researchers tried to infer many activity classes (basing themselves on the classes used in household travel surveys (HTSs)) and did not (gravely) aggregate activity classes (Simas-Oliveira et al. 2014). Others acknowledge the effect of the number of activity type classes and explicitly mention and/or experiment with this (Lu, Zhu, and Zhang 2012; Montini et al. 2014; Lu, Zhu, and Zhang 2013; Lu and Zhang 2015). However, none built a strong justification for the reason some activity types were aggregated, nor did they look at the effects of this aggregation other than increasing the classification accuracy. Yet, some researches may require pre-defined categories, which in that case over-rule the desire to use an optimised set of activity type classes. However, without such special constraints, the benefit of using an optimised set is more valuable than solely aiming at reaching a high classification accuracy.

Emerging activity type annotation and DF studies focus on the classification of each activity type. Special attention is devoted towards the activity-based approach. Various activity classes can be found in the literature, underlying the contextual or regional background. Vovsha et al. (2004) used three popular activity type classes regarding daily activity-travel patterns: mandatory (e.g. working, studying), non-mandatory (consisting of maintenance and discretionary activities), and at-home activities. This activity class categorisation scheme reflects the time and location constraints and is used in activity-based transport models (ABM).

Additionally, in rule-based ABM activity behaviour results from the rules of interaction between the preferences and constraints (Arentze et al. 2000). These models forecast daily schedules based on individual decision-making skeletons which reflect a variety of constraints, i.e. situational, institutional, household, spatial, temporal, and spatiotemporal constraints (Arentze and Timmermans 2004; Bellemans et al. 2010). Several studies also follow the original activity type classes from time-use surveys. Zhu et al. (2016) classified crowdsourcing data (Twitter text) based on the ten activity types from the American Time Use Survey (ATUS).

In Table 1, the distinct activity types that are used in Vovsha et al. (2004) are encoded based on the conditions of activity type, particularly work-related activities, and on time and location constraints.

[INSERT Table 1 HERE]

Activity classification methods

This section presents a summary of conventional data mining methods inferring activity types based on the behavioural attributes and their resources (see Table 2). The methods used in these studies can be classified into two major categories: the probability and the rule-based heuristic approaches. Each approach has different advantages and limitations regarding the classification process. Probability approach studies generally use the naïve Bayes classifier (Kusakabe & Asakura, 2014; Zhong et al., 2014) to estimate the probability of an alternative. In contrast, rule-based heuristic approach studies examine popular machine learning algorithms, e.g. decision trees (DTs), random forests (RFs), support vector machines (SVMs) or Bayesian belief networks (BBNs) to obtain the best classification accuracy for training (estimation) and test (validation) data sets (Toole et al.

2015; Reumers et al. 2013; Feng and Timmermans 2015).

The early stage studies, which often employ GPS-based data resources, recognised the importance of spatial information, i.e. point of interests (POIs). These studies pointed out that this location information can improve the activity type identification (Wolf et al. 2004; Bohte and Maat 2009; Stopher, FitzGerald, and Zhang 2008; Wolf, Guensler, and Bachman 2001). Recent activity type annotation studies also perceive the significance of additional information. Location and sequential information are explored in order to develop an innovative way to improve the activity type classification accuracy.

[INSERT Table 2 HERE]

Furthermore, the inferred activity types are utilised for validation (Liu et al. 2014), and the variability of activity-travel patterns in large data sets is investigated (Zhong et al., 2016).

An overview of the most popular classification algorithms (which are also used in this paper's comparison section) is provided next:

- C4.5 (J48): A decision tree learning algorithm which supports i.a. nominal and numeric attributes and offers advanced pruning capabilities. A minimum number of records for each leaf can be imposed. In this research it was set to 1.5% of the training set size.
- LMT: An ensemble classifier which builds a decision tree with a logistic model in each of the leafs. The same options were set as in the J48 algorithm.
- CHAID: A decision tree technique in which decisions are made based on chi-square and F-tests. 'CHAID' is an acronym for 'CHi-squared Automatic Interaction Detection'.

- Zero R: A reference classifier which predicts the most frequent activity type (a best-guess approach without taking into account any of the input attributes).
- One R: One single rule is used to predict activity type. This simple, lightweight classifier often yields a surprising classification accuracy.
- Random forest: An ensemble of decision tree learners. Each tree in the forest is built on a random vector of input data, sampled independently (and having the same distribution for all trees), and using a random selection of features (attributes). Similar options were used in this study as for the J48 decision tree learner.
- Logistic: A multinomial logistic regression model.
- Simple Logistic: A linear logistic regression model is build using LogitBoost with simple regression functions serving as base learners.
- Naïve Bayes: A simple, yet robust probabilistic classifier.
- Support Vector Machine (SMO): One of the most robust and accurate classifiers, yet resource intensive to build.

Given the problem of having to allocate an activity type to a stop in one's travels, machine learning techniques are likely to give the most accurate classification. A human researcher might develop some intuitive rules to allocate an activity type, yet machine learning techniques consider much more rules and will therefore in general produce a better prediction of the activity type. Some machine learning techniques do not really allow to inspect the rules that make up the model and are like a black box (i.e. neural networks, SVM…). However, the decision tree algorithms which are used throughout this research do allow to investigate and reflect upon the rules into the very detail.

Challenges

Behavioural data mining studies have the following limitations: the absence of a generally accepted (and used) activity type class definition (Ahern et al. 2013, 29), and the complexity of fusing different data sets (each having their own strengths) in order to enhance the data mining process. This section reviews the current research challenges. Furthermore, a means to overcome these issues is proposed and the behavioural data mining methods are updated accordingly.

Lack of activity categorisation standards

The existing literature on activity-travel data mining indicates that there are no clear standards for activity type classes, grounded by a theoretical background. The majority of researchers disregard the importance of activity categorisation; and thus merely rely on the travel survey design, which is designed from a contextual or cultural background. Few studies provide a thorough reasoning on the categorisation of activity types. In all cases, this task is left to the discretion of the researcher and is open for interpretation. As mentioned earlier, some researchers employ activity type classes that reflect time and location constraints, others simply use the activity categorisation taxonomy from time-use surveys.

Moreover, in the variety of activity type classes that is regularly employed, it is often the case that the similarity of activity patterns causes difficulties when classifying activities into certain categories. For example, in the case of non-mandatory activities, such as leisure, social and personal activities, the nature of these activities makes it tough to predict the proper activity type. In order to solve this problem, some researchers reorganise the activity type classes according to their presumed similar patterns. Kusakabe and Asakura (2014), for instance, integrated the leisure and business activities. In other studies, the problem is avoided by aggregating the activity types that are hard to classify (Alexander et al., 2015). However, this approach is merely a solution to enhance the activity inference estimation, while accepting the loss of important activity information, without even quantifying it.

This lack of activity categorisation standards may be addressed in two steps. First, the optimal number of activity type classes and their composition can be investigated using machine learning algorithms. The current research attempts to identify the best balance point between the improvement of model accuracy as a result from aggregating activity type classes, and the preservation of activity information from the original data sets. Second, data mining methods can be used to infer activity type information. In this study, temporal attributes are employed in the data mining methods, to obtain activity type information.

Complexity of fusing multiple data sets

In general, temporal attributes are selected as key aspects in order to impute the activity type (or trip purpose) in big transport data (Shen and Stopher 2013; Kusakabe and Asakura 2014). GPS-based movement records consist of precise spatial and temporal attributes, however the personal and household information are not available. The early stage data mining studies were restricted to solely using temporal attributes, such as activity start time and duration, to detect the activity type.

Several activity classes are difficult to distinguish based on temporal attributes only, which causes problems to accurately infer them. Additional variables are required to detect the type of activities. Recent studies focused on spatial information, in particular the incorporation of GIS (geographic information system) data, and crowdsourced resources (e.g. open street map). Feng & Timmermans (2014) report an overview of researches which adopt spatial information to recognise the type of activities, and propose a rule-based heuristic data mining method using temporal attributes associated with dummy values of spatial information, i.e. the existence of certain activity-related facilities in zonal area. Zhong et al. (2014) also propose an integrated approach, using the posterior probability of travel purposes at (bus) stops with temporal changes. This interpolated areal probability is based on the building footprint information (of buildings surrounding the trip destination) using the spatial analysis method of inverse distance weighting (IDW). These studies yield a better performance than prior ones.

The current study only uses temporal attributes from the HTS data set. The restriction to temporal attributes guarantees a compatibility of the method to as many big data sources as possible. The method is however identical when including other types of attributes. The aforementioned challenge is currently not the focus and may be addressed in future research.

Methodology

Data description

Two household travel surveys were used in this research: the Seoul HTS, and the Flanders (Belgium) HTS called OVG. The Seoul data set was used to develop the methodology. It consists of self-reported surveys from approximately 76,000 individuals. It was organised in the Seoul Metropolitan Area (SMA) in 2010 to collect household activity-travel data.

The activity type grouping methodology (see subsequent section) was also applied to the OVG 3.0-4.5 data set to investigate its performance in a different setting. The OVG data set contains single-day travel diaries (including weekends) enriched with individual and household socio-demographical information for approximately 17,300 individuals. The survey was conducted in multiple phases from 2007 until 2013. Of the 17,300 participants, approximately 13,200 conducted at least one trip.

The use of HTS-data, which strictly speaking only contains trip purposes and potentially not the full activity participation history, is valid within the research contexts as big transport data usually also has trip characteristics (e.g. smart card data). The trip-structured data was converted in activity-structured data for this research.

Grouping of activity types

As detailed before, the choice of available activity types as alternatives in the classifier strongly affects the classification accuracy. This research seeks to classify activity type as accurately as possible, yet without losing too much information by aggregating (i.e. grouping or combining of activity type classes). An optimisation strategy was designed to find the most optimal set of classes with respect to classification accuracy and limiting information loss.

The optimisation strategy consists out of three stages:

(1) Generating all possible combinations of classes of activity types

(2) Training and testing classifiers on data that was transformed according to the activity type class combinations of the previous step

(3) Finding the optimal set of activity type classes

The first step generates all possible combinations of classes of activities. This is a brute-force approach; all combinations, how unlikely or incompatible they might seem, are generated. In a particular class of activity types, the order of these activity types is unimportant. This limits the number of possible combinations. However, the number of combinations grows extremely quickly with increasing number of activity types. For example, at 10 activity types more than 117,000 different combinations of classes are possible. The home activity was excluded from this experiment. Typically the 'home' activity is quite easy to classify and it is not the focus of this research, which is optimising the classification of out-of-home activities which may be observed in (big) transport data. Machine learning algorithms have discovered several rules that will accurately label an activity to be a home activity, and this solely based on temporal attributes (activity duration and start time). Reumers et al. (2013) very minutely discuss these rules and the accuracies for each activity type class. The home activity is predicted with the highest true positive rate (even 96.5% on real-world independent data) and one of the best precisions (83.2% on the same real-world data). Additionally, the share of home activities is quite large. Its good classification capability therefore has a major effect on the overall classification accuracy; and this obscures suboptimal or bad classifications of out-of-home activities.

An efficient, two-stage algorithm was constructed to generate all possible combinations of classes. Firstly, all integer partitions of the number of activity types are generated. For example, the partitions of 4 activity types are:

- 1+1+1+1
- 2+1+1
- 2+2
- 3+1
- 4

These partitions may be taken as the number of activity types in a class (activity type class sizes). Secondly, all combinations of activity types into these classes are generated in a recursive method. Below the different combinations of classes in case of 4 activity types are given to illustrate what type of activity type classes are generated.

- ([1],[2],[3],[4]) (no aggregation)
- ([1, 2],[3],[4])
- ([1, 3],[2],[4])
- ([1, 4],[2],[3])
- ([2, 3],[1],[4])
- ([2, 4],[1],[3])
- ([3, 4],[1],[2])
- ([1, 2],[3, 4])
- ([1, 3],[2, 4])
- ([1, 4],[2, 3])
- ([1, 2, 3],[4])
- ([1, 2, 4],[3])
- ([1, 3, 4],[2])
- ([2, 3, 4],[1])
- ([1, 2, 3, 4]) (one group)

Using this method, only a few seconds were needed to calculate more than 117,000 unique sets of combinations of classes of activity types in case of 10 activity types.

The second step of the optimisation strategy builds a classifier on the data using the different combinations of classes from the previous step. J48 (C4.5, a rule-based heuristic method) and the Naïve Bayes algorithm (a probability approach) were considered in this step. J48 is a decision tree algorithm and Naïve Bayes is a robust probabilistic classifier. Both are, respectively, one of the most popular rule- and probability-based classifier types. The J48 algorithm is consistently (slightly) more accurate than Naïve Bayes on this data (see also Table 8), and is very fast to estimate. Therefore, this classifier was used in this second step, as more than 117,000 classifiers needed to be estimated. On a server equipped with two Intel Xeon E5-2643 v2 processors (running at approximately 80% capacity, i.e. 20 threads) this took roughly 30 hours.

The third step of the optimisation approach consists of finding the optimal set of activity type classes. One must acknowledge that 'optimal' is relative to the intended application. In the current research, a careful balance needs to be considered between classification accuracy and retaining sufficient information. For example, classifying activity type in classes of {Mandatory, Flexible, Discretionary} activity types might result in very high accuracies, yet the information in the activity type class is limited.

To perform this careful balancing, two measures were defined as part of the optimisation:

(1) *The test set classification accuracy*: this represents the classification accuracy which should be maximised. As will be detailed later on, the test set represents part of the input data (25%) which was not used to train the classifier, and hence can be used as an independent set to validate the classifier's accuracy.

(2) *The activity type entropy [bits]*: this quantifies the information retained in the set of activity type classes. This measure is sometimes also referred to as the Shannon diversity index as it quantifies the diversity between classes (Shannon (1948), as quoted in Manaugh and Kreider (2013)). It is calculated as $S = -\sum_i p_i \log_2(p_i)$ where p_i is the probability on class i. It is the average number of bits of information needed to store an activity type. This figure has its maximum at no combined activity types (10 distinct types) and its minimum of zero at a single class of activity types (all information is lost due to this aggregation). The activity type entropy, or retained information, should be maximised.

One now clearly sees the need for optimisation. Increasing the class size will lead to a higher classification accuracy, but also to a lower amount of activity information (lower activity type entropy).

In the optimisation the **test set accuracy improvement** and **entropy reduction** (both with respect to the reference case of no activity type aggregation) are respectively maximised and minimised. This optimisation is illustrated graphically in Figure 1. In this scatter plot, the red line may be moved upwards to isolate the data point with the most favourable characteristics in both test set accuracy improvement and entropy reduction limitation. In other words, the red line of equal utility is moved upwards to isolate the point with highest utility. The slope of the red line in this figure is calculated by $m = R_A/R_E$, where R_A is the range in test set accuracy improvement and R_E the range in entropy reduction. This formula assumes that both measures are equally important in the optimisation. This assumption will be challenged in a sensitivity analysis further down this paper. R_A is calculated by subtracting the minimum test set classification accuracy from the maximum test set classification accuracy. Similarly, R_E is defined as the difference between maximum entropy (no aggregation of activity types) and minimum entropy (excluding the case were all activity types belong to a single class and the entropy is equal to zero).

This optimisation may also be performed numerically by finding the activity type grouping combination for which

$$\mathrm{U} = \frac{A_i - A_0}{R_A} - \frac{E_0 - E_i}{R_E} \qquad (1)$$

is maximised. In this formula, A_i is the test set accuracy and E_i the activity type entropy of a particular combination of activity type classes i. A_0 and E_0 are, respectively, the test set accuracy and activity type entropy of the reference case of no activity type aggregation.

[INSERT Figure 1 HERE]

Comparison of classification algorithms

Several data mining algorithms are available to classify activity type classes. A very large number of algorithms has been made available in the Weka software (Hall et al. 2009). Apart from a convenient user interface, the power of the Weka software is available as a Java library. In this research, a Java script was developed which makes use of the Weka Java library.

To predict the activity type class, several algorithms were experimented with. Firstly, the records with a home activity were removed from the HTS Seoul data. As detailed before, typically these are quite easily classified and increase the classification accuracy (which masks the classifier's inaccuracy on other activities). Secondly, the data was split in a training set (75%) and a test set (25%). The records were randomly assigned to one of these subsets. This is a common practice. The training set is used to build the classification model, and the test set is used to independently evaluate the classification accuracy of the classifier on unseen data (a small validation). Thirdly, several of the most popular classification algorithms were used to predict the activity type class using temporal attributes. These algorithms were briefly introduced in the literature review section. The list of algorithms used in this comparison includes the C4.5 (J48), LMT, CHAID, Zero R, One R, random forest, logistic, simple logistic, Naïve Bayes and support vector machine (SMO) algorithms.

Analysis Results

Grouping of activity types

The above methodology to find an optimal combination of classes of activity types was performed on the Seoul HTS data set, considering only time-related attributes such as activity start time and duration.

Table 3 lists the 10 best ranked activity type class combinations, and some other interesting combinations to compare them with. The most optimal combination in Table 3 groups the activity types ‘bring/get’, ‘back home’, ‘shopping’, ‘leisure’ and ‘personal’. The test set accuracy increased by approximately 25% (to 79%) compared to the reference case, i.e. no aggregation of activity types. The accuracy increase comes at a cost of losing (on average) 0.886 bits of activity information compared to the reference set.

The activity types that were aggregated together cannot straightforwardly be categorised as being ‘flexible’, ‘discretionary’ or ‘non-mandatory’ activities, neither can they be referenced to as ‘short activities’. For example, a ‘bring/get’ activity has an obligation to a third party and may not be rescheduled that easily. It therefore is not a ‘flexible’ activity. The second ranked set of groups in Table 3 additionally includes ‘business’ in the aggregated activity type class, which is definitely not a ‘flexible’, ‘discretionary’ or ‘non-mandatory’ activity in the current interpretation. ‘Shopping’ is an activity that could have a long duration, objecting against the ‘short activities’ label for this aggregated class of activity types.

However, all of these aggregated activities have in common that they don’t have a very fixed time of the day to start. Although they could have obligations towards external parties and may not be rescheduled very easily, they could -in principle- be scheduled at any time of the day without a strong preference. This therefore is a class of ‘uncertain’, hard to classify activities. Figure 2 confirms this hypothesis. The activity types that are aggregated have the ‘flattest’ density curves for activity start time, meaning that the probability on these activity types is more equally spread during the day compared to the other activity types which have large peaks and are therefore easier to predict.

‘Business’ and ‘back to office’ also have a relatively ‘flat’ activity start time density curve but are not included in the class of ‘uncertain’ activity types. Perhaps there is still some factor which makes these easier to predict; or aggregating these as well majorly affects the entropy. ‘Back to office’ is relatively easy to predict in the afternoon. The other activity types show a reduced probability, while ‘back to office’ (as the only one) experiences a prominent increased probability in the afternoon. ‘Back to office’ distinguishes itself by having, on average, a longer duration. The disadvantage of losing information (entropy) when aggregating outweighs the gain in activity type classification accuracy, and therefore both activity types are not grouped in the top-ranked result of Table 3.

[INSERT Table 3 AND Table 4 HERE]

[INSERT Figure 2 HERE]

In previous researches, often activity types are categorised according to expert opinion as either ‘mandatory’, ‘maintenance’ or ‘discretionary’ (e.g. Bradley & Vovsha (2005)). Having only three activity type classes, relatively good classification accuracies may be reported, yet a considerable amount of activity information may be lost. The equivalent combination of activity type classes in the Seoul HTS data set, ([1, 2, 3, 4, 6, 7],[5, 8],[9, 10]), has a test set accuracy of 69.2% and corresponds to 1.259 bits of activity type information. Not only results this very popular set of classes in only a 14.6 percent point classification improvement compared to about 25 percent points in the most optimal case, it does so at a cost of losing 1.22 bits of information, which is also greater than that of the optimal case (0.886 bits). The expert opinion set of activity type classes is thus a considerably worse scheme than the one proposed in this paper. Considering an equal importance of accuracy gain and information retention, this scheme performs actually worse than the reference case of no aggregation due to the significant loss of information (negative value for U).

In the domain of activity-based modelling, activities are often characterised by a variety of constraints, i.e. situational, institutional, household, spatial, temporal, and spatiotemporal constraints (Kochan 2012; Arentze and Timmermans 2004; Bellemans et al. 2010). This line of thought may be applied in the classification of activity types into three categories: ‘work’, ‘non-work fixed’ and ‘non-work flexible’, as indicated in Table 1. As the classification criteria better matches those in the optimisation algorithm (temporal attributes), this scheme of classes offers a more beneficial combination of test model accuracy gain and information retention than the ‘mandatory’, ‘maintenance’ or ‘discretionary’ scheme, as can be observed in Table 3. Yet, this combination of activity type classes in the Seoul HTS data set ([2, 5, 7, 8, 9],[1, 4, 6],[3]) is inferior to the proposed most optimal set of classes. Although its test set classification accuracy is slightly greater by 1 percent point, there is a significant loss in information compared to the optimal case (only 1.356 bits retained instead of 1.592 bits in the optimal case).

An attempt was made to aggregate activity types in a similar manner as in previous studies, however a perfectly identical set of classes is impossible due to differences in the data and survey design. These are also listed in Table 3, where the aggregation scheme ranked 33rd is similar to that in S. Lee and Hickman (2014), the one ranked 298th similar to the one used in Shen and Stopher (2013) and the ones ranked 3159th and 62948th inspired by Lu and Zhang (2015). All except the one ranked 62948th, inspired by Lu and Zhang (2015), have a classification accuracy which is higher than the most optimal scheme. This comes at a cost of losing considerable information, i.e. activity inference diversity. This was the main critique formulated before: accuracy is potentially preferred above being able to infer more activity types. The scheme similar to the one ranked 62948th from Lu and Zhang (2015) performs worse on both the classification accuracy and information retention, compared to the most optimal scheme.

The methodology is not limited to the Seoul study area; it may be applied to other data sets without constraints. The only practical limitation is the number of activities that are considered for aggregation: too many initial activity types will result in an extremely large set of combinations that need to be processed.

As a validation of the proposed method, the aggregation of activity types was also optimised for the OVG 3.0-4.5 data set, the household travel survey of Flanders, Belgium. Only temporal attributes were used, similar to the conditions available for a big transport data annotation exercise. Again the 'home' activity type was excluded from the optimisation, which also ensures that the results may be compared to the optimisation in the Seoul HTS where the same approach was used.

In total, more than 21,000 sets of classes of activity types were generated based on the nine activity types in the OVG data set ('home' activity excluded). Table 5 lists the ten best ranked activity type class combinations, and some other interesting combinations to compare them with. The methodology proved that 'shopping' and 'bring/get someone/something' should be aggregated together, as well as 'visit someone' and 'relaxation, sport and culture'. Interestingly, the latter class is a mix of activity types traditionally labelled as 'mandatory' or 'non-work fixed', and 'discretionary' or 'non-work flexible’. It appears that these activity types have similar, uncertain temporal profiles as illustrated in Figure 3 based on activity start time.

Intuitively, the different classes make sense too. ‘Shopping’ and ‘bring/get someone/ something’ both imply the acquisition of goods (or possibly people as well in case of ‘bring/get’) and have no especially preferred time they should be executed. ‘Visit someone’ and ‘relaxation, sport and culture’ may also be relatable. Both are a form of personal entertainment (i.e. leisure), either within an individual or familial context and may have similar temporal profiles. Both are probably the most discretionary activity types. ‘Walk, tour, run’ is also of very discretionary nature, but may be more easier to predict because of its typically short duration. It should therefore indeed not be merged in the most optimal case.

Additionally, the activity type classes ‘mandatory’, ‘maintenance’ and ‘discretionary’ ([2, 3, 6, 8],[4, 10],[5, 7, 9]; ranked 8725th), as well as ‘work’, ‘non-work fixed’ and ‘non-work flexible’([3], [2, 6, 8],[4, 5, 7, 9, 10]; ranked 7171th) have been identified and included in Table 5. One can observe that these schemes once again perform considerably worse than the optimal activity aggregation scheme. In both schemes the test set classification accuracy is slightly higher than that of the optimal case (± 4 to 9 percent points), however at a significant cost: respectively only 1.564 and 1.370 bits of information have been retained from the original 3.010 bits. Yet, the most optimal combination of classes retains 2.431 bits of information.

Once more an attempt was made to aggregate activity types in a similar manner as in previous studies. These are listed in Table 5, where the scheme ranked 668th is similar to that in S. Lee and Hickman (2014), the one ranked 1001th similar to that in Lu, Zhu, and Zhang (2013), the schemes ranked 4455th, 11510th and 14231th similar to Lu and Zhang (2015) and the one ranked 7347th similar to that in Shen and Stopher (2013). With the exception of the scheme from Lu, Zhu, and Zhang (2013) and the one ranked 14231th from Lu and Zhang (2015), all have a higher test set classification accuracy than was found to be optimal, and hence at a cost of losing too much information. Again, accuracy appears to be preferred over being able to infer more activity types. The scheme similar to the one in Lu, Zhu, and Zhang (2013) and the one ranked 14231th from Lu and Zhang (2015) perform worse on both the classification accuracy and information retention, compared to the most optimal scheme.

[INSERT Table 5 AND Table 6 HERE]

[INSERT Figure 3 HERE]

Sensitivity analysis

It was mentioned before that the optimisation strategy assumes that both the test set accuracy improvement and the entropy reduction indices are equally important. This might not be true in all types of research. If there exists an intrinsic bias for one of the indices, a relative weight can be applied to Equation 1:

$$\mathrm{U} = \frac{A_i - A_0}{R_A} - a\,\frac{E_0 - E_i}{R_E} \tag{2}$$

In Equation 2, such a weight 'a' has been applied to the entropy reduction index. As both indices are normalised with the range in the denominator, a weight could be placed on either indices with the same result.

To see the effect of unequal importance of both indices, a small sensitivity analysis was performed in which weight 'a' was varied between 75% and 125% in steps of 1%. Figure 4 illustrates this process graphically. For each of the 51 resulting trials, the optimal combination of activity type classes was determined through maximisation of U. A frequency table was composed of these 51 sets of activity type classes, which resulted in Table 7.

Table 7 shows that in case of a weight of up to 1.25 on one of both indices, more than 50% of the cases the *original* optimal set of activity type classes will be put forward by the optimisation strategy. Figure 5 shows which combination of activity type classes was selected as most optimal in function of the weight ‘a’ on the entropy reduction index. A weight of 0% (a=1) corresponds to the case where the two indices are considered equally important. The ‘original’ most optimal set of activity type classes remains the most optimal for values of $a \in [0.97, 1.23]$. This optimum is relatively insensitive towards attributing a greater importance to the entropy reduction index. However, reducing the importance of this index by more than 3% will result in a different optimal set of activity type classes. Still, that particular set differs only by a single activity type that joined the class of aggregated activities. In fact, this set of classes is the second most optimal one in the unweighted optimisation of Table 3.

The influence of the weight on the entropy reduction index is quite consistent. If a small weight is applied on the entropy reduction index, i.e. attributing more importance to the test set accuracy improvement, larger classes of activity types are formed. If a larger weight is applied, i.e. rendering the test set accuracy index less important, one sees that the class size will reduce and more distinct activity types are retained. Note that these observations are based on the SMA dataset. Other study areas might experience a different sensitivity towards unequal weight situations.

[INSERT Table 7 HERE]

[INSERT Figure 4 and Figure 5 HERE]

Comparison of classification algorithms

The Weka (Hall et al. 2009) Java library was used to test several classification algorithms. A selection of algorithms was used to evaluate the effect of the different algorithms, and of the set of activity type classes (no aggregation, or the most optimal set of classes with {1,2,8,9,10}, as in Table 3). For this, the Seoul HTS data was used. Table 8 lists the classification accuracies on both the training and test set for these experiments, as well as an indication of their execution time. The different algorithms have similar relative accuracies in all the schemes. The absolute magnitude of the accuracies however differs between the different aggregation schemes, as expected.

One may observe that LMT consistently is ranked as the most accurate classifier on this data (based on the test set accuracy). However, it has to be remarked that training this classifier requires the most memory and processing time of all algorithms, except for the SMO algorithm which required almost twice the processing time of that of LMT.

J48, the algorithm of choice in the activity type class optimisation strategy, has overall the second best classification accuracy. It is however much faster compared to the more advanced LMT algorithm. The latter is similar to J48, but additionally trains a logistic classifier in each of the leafs of the decision tree (hence the improved accuracy, but considerably larger resource dependence). The unpruned version of this algorithm does not prune the branches of the decision tree in order to minimise classification errors and overfitting. This decision tree is considerably larger than the pruned version, but overfitting did not seem to occur.

The performance of the random forest algorithm is -overall- quite good. However, there seems to be a case of overfitting. The accuracy on the training set reaches 64.29%, whilst that of the test set is 53.27% (consistent with that of other algorithms). According to Breiman (2001), random forests should not overfit due to the Law of Large Numbers. The experiment was repeated with 100, 200 and 300 trees in the random forest, without significant improvements. Additionally, the number of randomly selected features to train each tree was varied as well, yet the apparent case of overfitting remained.

Remarkable is the performance of the One R algorithm. It is at least as good as several other, complex algorithms (sometimes even better than the support vector classifier). The rule created by the One R algorithm is based on activity duration.

The most optimal activity type combination strategy results in a test set accuracy increase of approximately 25 percent points, and this for all of the classification algorithms (except of course Zero R). The accuracy increase comes at a cost of losing (on average) 0.886 bits of activity information. However, this cost appears to be a property of the data set (no aggregation, or the most optimal set of activity type classes) and independent of the classifier. One therefore is free to choose the most optimal classifier. The LMT algorithm performs slightly better than J48, having a test set accuracy of 79.25% compared to 79.12% for J48. Nevertheless, this marginal accuracy increase does not justify the added cost in complexity and required resources.

[INSERT Table 8 HERE]

Discussion and Conclusion

Recent sources of data such as for example call detail records (CDR) or GPS enabled smartphone applications are incredibly valuable when processing spatiotemporal information. However, the reason why people are making trips (i.e. their trip purpose) is not known. This lacuna has spawned a whole new domain of trip purpose annotation (activity type classification, inference, imputation) researches. This study developed an optimised classification method for inferring the activity type from (big) transport data. The paper focuses on a rigorous activity type categorisation method (into aggregated activity type classes), something most studies disregarded. It demonstrates that most existing researches actually use a sub-optimal set of activity type classes in their methodology, leading to high classification accuracies, but low information in the prediction. The current research experimented with various machine learning algorithms as well.

The optimisation strategy for estimating the optimal set of activity type classes makes use of well-known classifier machine learning algorithms which follow the rule-based heuristic approach, and is composed of three stages. First, all possible combinations of classes of activity types are generated (home activities were excluded). Secondly, J48 (C4.5) classifiers are trained based on the Seoul HTS temporal attributes and the activity type class combinations in the previous step. Thirdly, the optimal set of activity type classes are identified, upholding the best balancing point between (test-set) accuracy improvements and retaining the activity type information (i.e. diversity by means of an activity entropy index). Similar temporal profiles (by means of kernel density distributions) were found for the activity types which were aggregated together by this staged approach. These aggregated activity types could however entail different functions and priorities (e.g. with or without obligation to a third party). Other variables than the currently used temporal variables, e.g. spatial indicators, could result in more disaggregated activity type classes (provided that these variables are good predictors for the activity type). This may be addressed in future research.

This method was also applied to the OVG data set (HTS of Flanders, Belgium) as a validation process of the methodology. Results seem to be similar to those based on the Seoul HTS. The result confirms that this method may be applied to all kinds of (HTS) data sets in order to distinguish the optimal activity type classes.

To investigate the effect of unequal importance of the classification accuracy improvement index, and the entropy reduction index, a small sensitivity analysis was performed.

Another contribution of current study is to compare different machine learning algorithms to perform the classification with. Popular data mining algorithms were tested using the Weka software Java libraries. The LMT algorithm ranked as the most accurate classifier in respect of test set accuracy, however it demands a considerable amount of memory and processing time. The decision tree classifier J48, Weka's implementation of C4.5, represented the second best classification accuracy. However because of faster processing time and the negligible accuracy reduction compared to LMT, it was used in the activity type class optimization. Noteworthy, the performance of One R algorithm is impressive; it is at least as good as several other, complex algorithms.

This study developed an optimised classification method towards the annotation of big transport data. The proposed method may be applied to distinguish the optimal activity type classes of any study area. The method may be used in AB models to optimise their schedule prediction components. Usually these models have troubles with accurately predicting non-mandatory activities (i.e. not work or education). The proposed optimisation strategy could assist in determining the best activity type classes used in such non-mandatory activity prediction sub-models. An additional advantage of this methodology is that it can automatically be applied to input data from any study area (having their own culture, different survey designs etc.) and use the output information to optimise activity type classification and schedule generation in a dedicated model for that study area.

The methodology considered each activity to be intrinsically equally important. Combining the entropy reduction index and a weighted classifier (such as a weighted decision tree) to determine the test set accuracy improvement index could provide an appropriate solution in case activity class weights are desired in a particular study or application.

Future research should focus on methods to develop more concrete classifiers using additional parameters, such as disaggregated location information.

Acknowledgements

This work was supported by the National Research Foundation of Korea funded by the Korean government (MSIP) under Grant NRF-2010-0028693.

References

Ahern, Aoife, Gill Weyman, Martin Redelbach, Angelika Schulz, Lars Akkermans, Lorenzo Vannacci, Eleni Anoyrkati, and Anouk Von grinsven. 2013. *Analysis of National Travel Statistics in Europe*. doi:10.2788/59474.

Alexander, Lauren, Shan Jiang, Mikel Murga, and Marta C. González. 2015. “Origin–destination Trips by Purpose and Time of Day Inferred from Mobile Phone Data.” *Transportation Research Part C: Emerging Technologies* 58 (September). Elsevier Ltd: 240–250. doi:10.1016/j.trc.2015.02.018.

Ali, Atizaz, Jooyoung Kim, and Seungjae Lee. 2016. “Travel Behavior Analysis Using Smart Card Data.” *KSCE Journal of Civil Engineering* 20 (4): 1532–1539. doi:10.1007/s12205-015-1694-0.

Allahviranloo, Mahdieh, and Will Recker. 2015. “Mining Activity Pattern Trajectories and Allocating Activities in the Network.” *Transportation* 42 (4). Springer US: 561–579. doi:10.1007/s11116-015-9602-5.

Arentze, Theo a., and Harry J.P. Timmermans. 2004. “A Learning-Based Transportation

Oriented Simulation System." *Transportation Research Part B: Methodological* 38 (7): 613–633. doi:10.1016/j.trb.2002.10.001.

Arentze, Theo, Frank Hofman, Henk van Mourik, and Harry J.P. Timmermans. 2000. "ALBATROSS: Multiagent, Rule-Based Model of Activity Pattern Decisions." *Transportation Research Record: Journal of the Transportation Research Board* 1706 (1): 136–144. doi:10.3141/1706-16.

Bellemans, Tom, Bruno Kochan, Davy Janssens, Geert Wets, Theo Arentze, and Harry J.P. Timmermans. 2010. "Implementation Framework and Development Trajectory of FEATHERS Activity-Based Simulation Platform." *Transportation Research Record: Journal of the Transportation Research Board* 2175 (1): 111–119. doi:10.3141/2175-13.

Bohte, Wendy, and Kees Maat. 2009. "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* 17 (3). Elsevier Ltd: 285–297. doi:10.1016/j.trc.2008.11.004.

Bradley, Mark, and Peter Vovsha. 2005. "A Model for Joint Choice of Daily Activity Pattern Types of Household Members." *Transportation* 32 (5): 545–571. doi:10.1007/s11116-005-5761-0.

Breiman, L. 2001. "Random Forests." *Machine Learning* 45 (1): 5–32. doi:10.1023/A:1010933404324.

Faouzi, Nour-Eddin El, Henry Leung, and Ajeesh Kurian. 2011. "Data Fusion in Intelligent Transportation Systems: Progress and Challenges – A Survey." *Information Fusion* 12 (1). Elsevier B.V.: 4–10. doi:10.1016/j.inffus.2010.06.001.

Feng, Tao, and Harry J.P. Timmermans. 2014. "Extracting Activity-Travel Diaries from GPS Data: Towards Integrated Semi-Automatic Imputation." *Procedia Environmental Sciences* 22. Elsevier B.V.: 178–185. doi:10.1016/j.proenv.2014.11.018.

Feng, Tao, and Harry J.P. Timmermans. 2015. "Detecting Activity Type from GPS Traces Using Spatial and Temporal Information." *European Journal of Transport and Infrastructure Research* 15 (4): 662–674.

Gong, Lei, Takayuki Morikawa, Toshiyuki Yamamoto, and Hitomi Sato. 2014. "Deriving Personal Trip Data from GPS Data: A Literature Review on the Existing

Methodologies.” *Procedia - Social and Behavioral Sciences* 138 (June): 557–565. doi:10.1016/j.sbspro.2014.07.239.

Graham, M., and T. Shelton. 2013. “Geography and the Future of Big Data, Big Data and the Future of Geography.” *Dialogues in Human Geography* 3 (3): 255–261. doi:10.1177/2043820613513121.

Hall, Mark, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, and Ian H Witten. 2009. “The WEKA Data Mining Software.” *ACM SIGKDD Explorations* 11 (1): 10–18. doi:10.1145/1656274.1656278.

Järv, Olle, Rein Ahas, and Frank Witlox. 2014. “Understanding Monthly Variability in Human Activity Spaces: A Twelve-Month Study Using Mobile Phone Call Detail Records.” *Transportation Research Part C: Emerging Technologies* 38 (January). Elsevier Ltd: 122–135. doi:10.1016/j.trc.2013.11.003.

Kitchin, R. 2013. “Big Data and Human Geography: Opportunities, Challenges and Risks.” *Dialogues in Human Geography* 3 (3): 262–267. doi:10.1177/2043820613513388.

Kochan, Bruno. 2012. “Implementation, Validation and Application of an Activity-Based Transportation Model for Flanders.” https://uhdspace.uhasselt.be/dspace/bitstream/1942/15067/1/PhD thesis Bruno Kochan.pdf.

Kusakabe, Takahiko, and Yasuo Asakura. 2014. “Behavioural Data Mining of Transit Smart Card Data: A Data Fusion Approach.” *Transportation Research Part C: Emerging Technologies* 46 (September). Elsevier Ltd: 179–191. doi:10.1016/j.trc.2014.05.012.

Lee, SangGu, and Mark Hickman. 2014. “Trip Purpose Inference Using Automated Fare Collection Data.” *Public Transport* 6 (1–2): 1–20. doi:10.1007/s12469-013-0077-5.

Lee, Won Do, Keechoo Choi, Tom Bellemans, Jeong Hwan Hwang, and Sungjin Cho. 2015. “Data Mining Method for Smart Card Data Using Household Travel Survey: A Pilot Study of Public Transportation in Suwon, South Korea.” In *International Association of Traveler Behavior Conference*.

Liu, Feng, Davy Janssens, JianXun Cui, YunPeng Wang, Geert Wets, and Mario Cools. 2014. “Building a Validation Measure for Activity-Based Transportation Models Based on Mobile Phone Data.” *Expert Systems with Applications* 41 (14). Elsevier

Ltd: 6174–6189. doi:10.1016/j.eswa.2014.03.054.

Liu, Feng, Davy Janssens, Geert Wets, and Mario Cools. 2013. “Annotating Mobile Phone Location Data with Activity Purposes Using Machine Learning Algorithms.” *Expert Systems with Applications* 40 (8): 3299–3311. doi:10.1016/j.eswa.2012.12.100.

Lu, Yijing, and Lei Zhang. 2015. “Imputing Trip Purposes for Long-Distance Travel.” *Transportation* 42 (4). Springer US: 581–595. doi:10.1007/s11116-015-9595-0.

Lu, Yijing, Shanjiang Zhu, and Lei Zhang. 2012. “A Machine Learning Approach to Trip Purpose Imputation in GPS-Based Travel Surveys.” In *Presented at 4th Conference on Innovations in Travel Modeling, Tampa, May 2012*, 1–8.

Lu, Yijing, Shanjiang Zhu, and Lei Zhang. 2013. “Imputing Trip Purpose Based on GPS Travel Survey Data and Machine Learning Methods.” *Transportation Research Record* 1250.

Manaugh, Kevin, and Tyler Kreider. 2013. “What Is Mixed Use? Presenting an Interaction Method for Measuring Land Use Mix.” *Journal of Transport and Land Use* 6 (1): 63–72. doi:10.5198/jtlu.v6i1.291.

Montini, Lara, Nadine Rieser-Schüssler, Andreas Horni, and Kay W. Axhausen. 2014. “Trip Purpose Identification from GPS Tracks.” *Transportation Research Record: Journal of the Transportation Research Board* 2405: 16–23. doi:10.3141/2405-03.

Nurul Habib, Khandker M., and Eric J. Miller. 2009. “Modelling Activity Generation: A Utility-Based Model for Activity-Agenda Formation.” *Transportmetrica* 5 (1): 3–23. doi:10.1080/18128600802591384.

Reumers, Sofie, Feng Liu, Davy Janssens, Mario Cools, and Geert Wets. 2013. “Semantic Annotation of Global Positioning System Traces.” *Transportation Research Record: Journal of the Transportation Research Board* 2383 (December): 35–43. doi:10.3141/2383-05.

Reumers, Sofie, Feng Liu, Davy Janssens, and Geert Wets. 2014. “The Annotation of Global Positioning System (GPS) Data with Activity Purposes Using Multiple Machine Learning Algorithms.” In *Mobile Technologies for Activity-Travel Data Collection and Analysis*, edited by Soora Rasouli and Harry Timmermans, i:119–133. IGI Global. doi:10.4018/978-1-4666-6170-7.ch008.

Shannon, C. E. 1948. “A Mathematical Theory of Communication.” *Bell System Technical*

Journal 27 (3). Alcatel-Lucent: 379–423. doi:10.1002/j.1538-7305.1948.tb01338.x.

Shen, Li, and Peter R. Stopher. 2013. “A Process for Trip Purpose Imputation from Global Positioning System Data.” *Transportation Research Part C: Emerging Technologies* 36 (November): 261–267. doi:10.1016/j.trc.2013.09.004.

Simas-Oliveira, Marcelo G., Peter Vovsha, Jean Wolf, and Michael Mitchell. 2014. “Evaluating Two Methods for Identifying Trip Purpose in GPS-Based Household Travel Surveys Evaluating Two Methods for Identifying Trip Purpose in GPS-Based Household Travel Surveys.” *Transportation Research Record* 2405: 33–41. doi:10.3141/2405-05.

Stopher, Peter R., Camden FitzGerald, and Jun Zhang. 2008. “Search for a Global Positioning System Device to Measure Person Travel.” *Transportation Research Part C: Emerging Technologies* 16 (3): 350–369. doi:10.1016/j.trc.2007.10.002.

Toole, Jameson L., Serdar Colak, Bradley Sturt, Lauren P. Alexander, Alexandre Evsukoff, and Marta C. González. 2015. “The Path Most Traveled: Travel Demand Estimation Using Big Data Resources.” *Transportation Research Part C: Emerging Technologies* 58. Elsevier Ltd: 162–177. doi:10.1016/j.trc.2015.04.022.

Vovsha, Peter, Eric Petersen, and Robert Donnelly. 2004. “Impact of Intrahousehold Interactions on Individual Daily Activity-Travel Patterns.” *Transportation Research Record* 1898 (1898): 87–97. doi:10.3141/1898-11.

Wang, Zhenzhen, Sylvia Y. He, and Yee Leung. 2017. “Applying Mobile Phone Data to Travel Behaviour Research: A Literature Review.” *Travel Behaviour and Society*. Hong Kong Society for Transportation Studies. doi:10.1016/j.tbs.2017.02.005.

Wolf, Jean, Randall Guensler, and William Bachman. 2001. “Elimination of the Travel Diary: Experiment to Derive Trip Purpose from Global Positioning System Travel Data.” *Transportation Research Record* 1768 (1): 125–134. doi:10.3141/1768-15.

Wolf, Jean, S. Schönfelder, U. Samaga, M. Oliveira, and Kay Axhausen. 2004. “Eighty Weeks of Global Positioning System Traces: Approaches to Enriching Trip Information.” *Transportation Research Record: Journal of the Transportation Research Board* 1870 (August): 46–54. doi:10.3141/1870-06.

Yue, Yang, Tian Lan, Anthony G.O. Yeh, and Qing-Quan Li. 2014. “Zooming into Individuals to Understand the Collective: A Review of Trajectory-Based Travel

Behaviour Studies." *Travel Behaviour and Society* 1 (2). Hong Kong Society for Transportation Studies: 69–78. doi:10.1016/j.tbs.2013.12.002.

Zhao, Fang, Francisco Câmara Pereira, Rudi Ball, Youngsung Kim, Yafei Han, Christopher Zegras, and Moshe Ben-Akiva. 2015. "Exploratory Analysis of a Smartphone-Based Travel Survey in Singapore." *Transportation Research Record: Journal of the Transportation Research Board* 2 (December): 45–56. doi:10.3141/2494-06.

Zhong, Chen, Michael Batty, Ed Manley, Jiaqiu Wang, Zijia Wang, Feng Chen, and Gerhard Schmitt. 2016. "Variability in Regularity: Mining Temporal Mobility Patterns in London, Singapore and Beijing Using Smart-Card Data." Edited by Tobias Preis. *PLOS ONE* 11 (2): e0149222. doi:10.1371/journal.pone.0149222.

Zhong, Chen, Xianfeng Huang, Stefan Müller Arisona, Gerhard Schmitt, and Michael Batty. 2014. "Inferring Building Functions from a Probabilistic Model Using Public Transportation Data." *Computers, Environment and Urban Systems* 48 (November). Elsevier Ltd: 124–137. doi:10.1016/j.compenvurbsys.2014.07.004.

Zhu, Zack, Ulf Blanke, and Gerhard Tröster. 2016. "Recognizing Composite Daily Activities from Crowd-Labelled Social Media Data." *Pervasive and Mobile Computing* 26 (February). Elsevier B.V.: 103–120. doi:10.1016/j.pmcj.2015.10.007.

List of Tables

Table 1. Example of activity class encodings using time and location constraints.

Activity encoding of Vovsha et al. (2004)	Activity classes in Kochan (2012)	ABM cases		Constraints	
		ALBATROSS[a]	FEATHERS[b]	Time	Location
Mandatory (Md)	Work (W)	Work and school	Work	Fixed start & end time	Fixed destination
		Voluntary work	-	Fixed start & end time	Fixed destination
	Non-Work Fixed (NFi)	-	Business	Fixed start & end time	Fixed destination
		Bring/get	Bring/get	Fixed start & end time	Fixed destination
		In-home	Being at home	Unconstrained / Fixed start & end time	Home location
Maintenance (M)	Non-Work Flexible (NFl)	Daily shopping	Daily shopping	Opening hours	Available facilities
		Non-daily shopping	Non-daily shopping	Opening hours	Available facilities
		Services	Service	Opening hours	Available facilities
Discretionary (D)	Non-Work Flexible (NFl)	Social contacts (Out-of-home)	Social visits	Unconstrained	Every zone
		Leisure (Out-of-home)	Leisure	Opening hours	Available facilities
	Non-Work Fixed (NFi)	Other (Out-of-home)	Others	Fixed start & end time	Fixed destination

[a] ALBATROSS (A Learning-based Transportation Oriented Simulation System) was developed for the Dutch Ministry of Transportation, Public Works and Water Management, to explore possibilities of a rule-based approach and to develop a travel demand model for policies. For details of the model specification and information, please check Arentze & Timmermans (2004).

[b] FEATHERS (Forecasting Evolutionary Activity-Travel of Households and their Environmental RepercussionS) was developed by executing a research program coordinated by IMOB (Transportation Research Institute) in 2005, funded by IWT, in order to facilitate the development of a dynamic activity-based transportation demand model for Flanders, Belgium (Bellemans et al. 2010).

Table 2: Summary of recent data mining methods inferring activity types.

Paper	Purpose	Reference data sets	Target data sets	Approach	Data collection period	Study area
Reumers et al. (2013)	Inferring the activity types	Survey data	GPS movement data	Rule-based heuristic	2006-2007	Flanders, Belgium
Reumers et al. (2014)					2006-2007	Flanders, Belgium
Feng & Timmermans (2015)					2012-2013	Rotterdam regions, Netherlands
W. Do Lee et al. (2015)					Survey: 2009-2010	Seoul, South Korea
					Smart-card: 2011	
Kusakabe & Asakura (2014)			Smart-card data (public transits)	Probability	2007-2009	Osaka metropolitan area, Japan
Zhong et al. (2014)					Survey: 2008	Singapore
					Smart-card: 2011	
Allahviranloo & Recker (2015)			GPS movement data		2000-2001	California, U.S
Liu et al. (2013)	Underlying activity-travel patterns	Mobile phone call detail records (CDRs)		Rule-based heuristic	2009-2011	Flanders, Belgium
Ali et al. (2016)		Smart-card data (public transits)			2012	Seoul, South Korea
Liu et al. (2014)	Applications	Survey data	CDRs	Rule-based heuristic	Survey: 2003, 2006-2007	Survey: South Africa & Belgium
					CDRs: 2011	CDRs: Ivory Coast
Zhong et al. (2016)		Smart-card data		Probability	2013-2014	London, UK. Singapore Beijing, China

Table 3. Grouping activity types based on the temporal attributes (10 best ranked) in the Seoul HTS survey. Table 4 lists the activity type encoding.

Rank	**G1**	**G2**	**G3**	**G4**	**G5**	**G6**	**# groups**	**Null-model accuracy**	**Train model accuracy**	**Test model accuracy**	**Entropy [bits]**	**U**
1	1, 2, 8, 9, 10	3	4	5	6	7	6	0.395	0.781	0.791	1.592	0.172
2	1, 2, 6, 8, 9, 10	3	4	5	7		5	0.452	0.834	0.843	1.308	0.167
3	1, 2, 5, 8, 9, 10	3	4	6	7		5	0.413	0.800	0.812	1.467	0.165
4	1, 2, 8, 9, 10	4, 5	3	6	7		5	0.395	0.780	0.791	1.573	0.164
5	1, 2, 8, 9, 10	4, 7	3	5	6		5	0.395	0.781	0.791	1.572	0.164
6	1, 2, 7, 8, 9, 10	3	4	5	6		5	0.415	0.801	0.811	1.458	0.160
7	1, 2, 6, 8, 9, 10	4, 5	3	7			4	0.452	0.834	0.843	1.289	0.160
8	1, 2, 8, 9, 10	3, 4	5	6	7		5	0.400	0.784	0.793	1.550	0.159
9	1, 2, 8, 9, 10	4, 6	3	5	7		5	0.396	0.780	0.791	1.564	0.159
10	1, 2, 6, 8, 9, 10	4, 7	3	5			4	0.452	0.834	0.843	1.288	0.172
33[a]	1, 2, 5, 6, 7, 8, 9, 10	3	4				3	0.498	0.874	0.884	1.038	0.146
298[b]	1, 2, 5, 6, 7, 9, 10	8	3	4			4	0.423	0.794	0.806	1.371	0.112
536[c]	2, 5, 7, 8, 9, 10 (NFl)	1, 4, 6 (NFi)	3 (W)				3	0.419	0.801	0.802	1.356	0.099
3159[d]	1, 2, 4, 5, 7, 8, 9, 10	3, 6					2	0.501	0.835	0.844	0.998	0.043
62948[d]	1, 2, 4, 5, 7, 10	3, 6	8, 9				3	0.397	0.682	0.681	1.448	-0.128
95395[e]	1, 2, 3, 4, 6, 7 (Md)	5, 8 (M)	9, 10 (D)				3	0.484	0.694	0.692	1.259	-0.181
Ref.	No grouping						10	0.262	0.561	0.546	2.478	0

*Remark: type 0 (being at home) was excluded in this classification since this is typically well-classified and this research focuses on out-of-home activities, additionally significantly reducing computational intensity.

** Note: activity type classes listed are similar to those of previous studies and are based on:

[a] S. Lee and Hickman (2014)

[b] Shen and Stopher (2013)

[c] Arentze and Timmermans (2004), Bellemans et al. (2010), Kochan (2012)

[d] Lu and Zhang (2015)

[e] e.g. Bradley and Vovsha (2005)

Table 4: Activity type encoding in the Seoul HTS survey.

Trip purpose	**Description**	**Encoding based on Vovsha et al. (2004)**	**Encoding based on Kochan (2012)**
0	Being at home	-	-
1	Bring/get	Mandatory	Non-work fixed
2	Back home*	Mandatory	Non-work flexible
3	Work	Mandatory	Work
4	School	Mandatory	Non-work fixed
5	Education service	Maintenance	Non-work flexible
6	Business	Mandatory	Non-work fixed
7	Back to office*	Mandatory	Non-work flexible
8	Shopping	Maintenance	Non-work flexible
9	Leisure/recreation/communication	Discretionary	Non-work flexible
10	Personal & religious activities	Discretionary	Non-work flexible

*These are special activity types in the Seoul HTS, representing the actual act of travelling.

Table 5: Grouping activity types based on the temporal attributes (10 best ranked) in OVG 3.0-4.5. Table 6 lists the activity type encoding.

Rank	**G1**	**G2**	**G3**	**G4**	**G5**	**G6**	**G7**	**G8**	**# groups**	**Null-model accuracy**	**Train model accuracy**	**Test model accuracy**	**Entropy [bits]**	**U**
1	4, 8	5, 9	2	3	6	7	10		7	0.223	0.596	0.616	2.431	0.0871
2	4, 8	5, 9	3, 6	2	7	10			6	0.244	0.628	0.655	2.237	0.0867
3	4, 8	5, 7, 9	2	3	6	10			6	0.248	0.620	0.641	2.243	0.063
4	4, 8, 10	5, 9	2	3	6	7			6	0.262	0.638	0.650	2.197	0.062
5	4, 8, 10	5, 9	3, 6	2	7				5	0.282	0.670	0.687	2.003	0.059
6	4, 8	2	3	5	6	7	9	10	8	0.189	0.524	0.548	2.692	0.058
7	2, 4, 8	5, 9	3	6	7	10			6	0.276	0.637	0.656	2.144	0.055
8	4, 8	5, 7, 9	3, 6	2	10				5	0.268	0.651	0.675	2.049	0.054
9	4, 8	3, 6	2	5	7	9	10		7	0.210	0.558	0.584	2.498	0.053
10	4, 8	5, 9	2, 7	3	6	10			6	0.230	0.597	0.621	2.314	0.053
668[a]	4, 5, 7, 8, 9, 10	2, 3	6						3	0.554	0.816	0.814	1.103	-0.039
1001[b]	5, 7, 9	2, 3	4	6	8	10			6	0.218	0.555	0.556	2.351	-0.105
4455[c]	4, 5, 6, 7, 8, 9, 10	2, 3							2	0.648	0.836	0.834	0.775	-0.122
7171[d]	4, 5, 7, 9, 10	2, 6, 8	3						3	0.436	0.707	0.698	1.370	-0.153
7347[e]	5, 7, 8, 9, 10	2, 3	4	6					4	0.339	0.624	0.623	1.738	-0.155
8725[f]	2, 3, 6, 8	4, 10	5, 7, 9						3	0.343	0.644	0.650	1.564	-0.168
11510[c]	4, 5, 7, 9	6, 8, 10	2, 3						3	0.392	0.654	0.655	1.467	-0.194
14231[c]	4, 7, 9	6, 8, 10	2, 3	5					4	0.290	0.554	0.558	1.884	-0.219
Ref.	No grouping								9	0.136	0.445	0.452	3.010	0

*Remark: type 1 (home) was excluded in this classification since this is typically well-classified and this research focuses on out-of-home activities, additionally significantly reducing computational intensity.

** Note: activity type classes listed are *similar* to those of previous studies and are based on:

[a] S. Lee and Hickman (2014)

[b] Lu, Zhu, and Zhang (2013)

[c] Lu and Zhang (2015)

[d] Arentze and Timmermans (2004), Bellemans et al. (2010), Kochan (2012)

[e] Shen and Stopher (2013)

[f] e.g. Bradley and Vovsha (2005)

Table 6: Activity encoding in OVG 3.0-4.5 (the household travel survey from Flanders, Belgium).

Trip purpose	**Description**	**Encoding based on Vovsha et al. (2004)**	**Encoding based on Kochan (2012)**
1	Home	-	-
2	Business	Mandatory	Non-work fixed
3	Work	Mandatory	Work
4	Shopping	Maintenance	Non-work flexible
5	Visit someone	Discretionary	Non-work flexible
6	Education	Mandatory	Non-work fixed
7	Walk, tour, run	Discretionary	Non-work flexible
8	Bring/get someone/something	Mandatory	Non-work fixed
9	Relaxation, sport, culture	Discretionary	Non-work flexible
10	Services (physician, bank…)	Maintenance	Non-work flexible

Table 7: Result of the sensitivity analysis for the optimisation strategy on the SMA data. The relative weight of the 'test set accuracy improvement' and 'entropy reduction' indices was varied between 75% and 125% in steps of 1% (51 trials). The table lists the frequencies of *the most optimal activity type grouping* in each of those 51 trials. Figure 4 shows this process graphically.

'Optimised' activity type groups	**Frequency**	**Percent**	**Cumulative Percent**
([1, 2, 8, 9, 10], [3], [4], [5], [6], [7])	27	52.94%	52.94%
([1, 2, 6, 8, 9, 10], [3], [4], [5], [7])	12	23.53%	76.47%
([1, 2, 5, 6, 8, 9, 10], [3], [4], [7])	7	13.73%	90.20%
([1, 2, 5, 6, 7, 8, 9, 10], [3], [4])	3	5.88%	96.08%
([1, 8, 9, 10], [2], [3], [4], [5], [6], [7])	2	3.92%	100.00%
	51	100%	

Table 8: Comparison of different classification algorithms and sets of activity type classes (before and after optimisation). Training set (75%) and test set (25%) classification accuracies are listed. Runtime information on a workstation with 2x Intel Xeon (X5670) CPUs and 48GB RAM is provided (Random Forest estimation running on 12 threads, others single thread).

	No grouping			**Grouped {1,2,8,9,10}**		
	Train set	**Test set**	**Runtime [s]**	**Train set**	**Test set ↓**	**Runtime [s]**
LMT	57.32%	55.86%	202	78.29%	79.25%	121
J48	56.06%	54.60%	<5	78.07%	79.12%	<5
J48 (unpruned)	56.03%	54.52%	<5	78.07%	79.12%	<5
SMO (support vector classifier)	54.03%	52.45%	63	77.27%	78.36%	84
Simple Logistic	54.07%	52.38%	94	77.23%	78.34%	70
Logistic	53.91%	52.31%	60	77.18%	78.29%	39
One R	54.09%	52.70%	<5	77.11%	78.09%	<5
CHAID	56.00%	55.50%	<5	78.30%	77.90%	<5
Naïve Bayes	53.46%	51.54%	<5	76.67%	77.68%	<5
Random Forest	64.29%	53.27%	52	81.24%	77.53%	54
Zero R	45.70%	43.02%	<5	45.70%	43.02%	<5

List of Figures

Figure 1: Illustration of Step 3 in the activity type grouping optimisation.

Figure 2: Kernel density plots of activity start time in Seoul HTS for the best-ranked activity type grouping set from Table 3.

Figure 3: Kernel density plots of activity start time in OVG 3.0-4.5 (Flanders, Belgium) for the best-ranked activity type grouping set from Table 5.

Figure 4: Illustration of the sensitivity analysis methodology for the optimisation strategy on the SMA data. The slope of the line of equal utility was altered between 75% and 125% in steps of 1%. The result of this sensitivity analysis is listed in Table 7.

Figure 5: The *most optimal combination of activity type classes* in function of the weight imposed on the *entropy reduction* index. A weight of -25% corresponds to a=0.75 and a weight of +25% to a=1.25.