Vehicle Availability Profiling from Diverse Data Sources

Sophie Naylor
Dept. of Architecture & Built Environment
University of Nottingham
Nottingham, UK
sophie.naylor@nottingham.ac.uk

James Pinchin
Horizon Digital Economy Research
Dept. of Architecture & Built Environment
University of Nottingham
Nottingham, UK

Rebecca Gough
Helios
A. T. Kearney
London, UK

Mark Gillott
Dept. of Architecture & Built Environment
University of Nottingham
Nottingham, UK

An understanding of the movement and utilisation rates of vehicles has many applications in the 'smart city'. The increasing availability of location and movement data from smartphones, in-vehicle loggers etc. allows for new applications of vehicle use analytics to be developed using a diverse range of data sources. However, for wide-scale application it may not be feasible to rely on consistent data from all possible sources - a need arises to process disparate data sources into a unified format. Raw user location data also presents significant issues around user privacy and the need to securely store and transmit any personally identifiable data. This paper covers the problem definition and development of a system to classify vehicle user driving patterns. A system was proposed to allow user driving patterns to be characterised in a way that does not explicitly store large volumes of location data while retaining key information needed for behavioural analysis. User driving data was converted to a personal profile based on statistical likelihood of vehicle use over a 24-hour period. Dynamic Time Warping was used to quantify the match between a new user's calculated profile and established driving archetypes. Additional profile features were tested in trained multi-class classification models including typical journey length, no. journeys etc. This was found to reduce the number of days of data needed to make a match for most users. This increases the feasibility of representing vehicle users in relation to driving archetypes rather than explicitly storing sensitive location data.

Keywords— vehicle tracking, location data, activity profiles

I. Introduction

With the increasing availability of wide-scale location tracking and behavioral data capture, the location patterns of people and tracked assets have seen an increasing variety of uses in recent years. A broad range of services and applications provide personalized services, analyze trends over large populations or optimize systems in real time based on location tracking data. The field of vehicle tracking and vehicle-based behavior analysis is no exception to this trend, with vehicle location and usage patterns of particular interest to a range of applications, including:

- Navigation and optimization services – while GNSS-based navigation services are an established field, route optimization services can require wider ranges of data from other vehicles etc.
- Fleet management – commercial fleet management packages can include real-time position monitoring of fleet vehicles for driver behavior tracking or route optimization [1]. Location monitoring can also be used to ensure appropriate asset utilization rates and improve fleet efficiency.
- Car park management – monitoring and optimization of parking space utilization rates require a localized measure of vehicle use patterns.
- Smart grid & Vehicle-to-grid services – for electric vehicles, use of the on-board vehicle battery for smart grid optimization can require localization and knowledge of vehicle use patterns. Driver behavioral patterns have been shown to affect the potential energy impact of V2G services [2].
- Smart city applications – for wider analyses of people flows, services etc. within a smart city environment, vehicle mobility data can provide value as a data stream.
- Internet of Things – vehicles are expected to be further integrated into wider Internet of Things services [3], potentially requiring location and use to be monitored as part of smart service offering.

Many of the above applications have the common thread of tracking the vehicle as an asset to be utilized in some form. Applications in this vein often require the availability of the tracked assets to be measured, characterized and in some cases predicted. In the case of vehicle tracking, this introduces the need to understand the behavioral patterns of the vehicle user.

The application of positioning technology to the tracking of vehicles at scale, especially where centralised processing of data from multiple vehicles is required, introduces issues around how data can be collected to ensure clarity, consistency, reliability and to maintain privacy of the vehicle user. This paper seeks a solution to allow vehicle usage patterns to be characterised for behavioural analysis or prediction, combining data from multiple sources that include location-based and non-location based data.

II. Vehicle Data Issues

The range of data collection strategies for vehicle location tracking present some issues for wide-scale application. Most data collection methods rely on the introduction of additional sensor hardware or software to each monitored vehicle, which makes consistent application

Funding for this research provided by Innovate UK and A.T. Kearney through Innovate UK project number KTP10855.

of a single solution across diverse vehicle and user types less feasible. In wide-scale application, it must be possible to process and compare vehicle tracking data that may arrive in different formats.

An additional issue arises in user-centred vehicle tracking. Where a single user drives a single vehicle, the issue of tracking user or vehicle does not present any major reliability problems. However, a significant number of vehicles have multiple potential drivers, both in domestic contexts (family vehicles) and commercial (fleet applications). In this case, relying on a single user as data source for a multi-user vehicle may introduce false negative errors into any application. In applications that rely on consistent knowledge of vehicle status such as vehicle-to-grid battery use, this can introduce potential failures in the system.

As with any application collecting location data from end users, vehicle tracking also presents challenges related to the privacy and data security of its participants, especially in light of recent data policy changes such as the EU General Data Protection Regulation (GDPR) [4]. Location data in particular is considered highly sensitive, with as few as four location points making it possible to uniquely identify individuals [5]. Increasing awareness of the issues around data privacy and its implications in data-analysis applications are reflected in recent policy changes [6]. It is essential from both an ethical and legal standpoint that sensitive data, whether collected directly or inferred through combination of sources, is minimised where possible and treated appropriately where collected.

It is therefore necessary to consider data collection and analysis that is as robust as possible to uncertainties, while also minimising the amount of sensitive user data that is collected, transmitted and stored.

III. Methodology

The methods presented in this study seek to address the data collection and processing issues identified in the previous sections: to extract vehicle use profiles from collected user data and represent them in a more abstracted form. For the purpose of behavioural pattern profiling, statistical profiles of vehicle use were constructed. A major focus of this work was the 'activity profile', where user driving data was converted to a personal profile of the statistical likelihood of vehicle use over a 24h period, separated by weekday/weekend. This statistical profile offers the following benefits:

- Allows the clustering of similar users by a typical pattern of activity.
- Allows further analysis to be a step removed from the variations caused by different data sources.
- Addresses some privacy issues by removing direct location data associated with the user.
- Potentially allows users to be represented as a combination of statistics rather than storing specific user locations over longer time periods.

It was proposed that domestic vehicle users could be categorised against a set of driver archetypes representing common driving behavioural patterns. These archetypes were represented as a set of baseline activity profiles to which new user data could be compared. The following sections of this paper present the development of vehicle use archetypes and the classification of new vehicle profiles.

The generation of the vehicle use archetypes was achieved using an Agent-Based simulation model to generate large volumes of driving data according to predefined rules and is presented in Section V. This model was used to create controllable 'baseline' versions of the defined user archetypes, as well as generate large datasets of simulated users with a known archetype for testing. The classification of new users was achieved using trained machine learning models, presented in Section VI.

IV. Data Sources Used

The vehicle activity profile was designed to be constructed from diverse data sources. In this study, three data types were used for testing: vehicle-based GPS/speed, simulated activity data and user-based personal device data.

A. GPS/Speed Data

The major dataset used in this study is a publicly available dataset of electric vehicle drive cycles collected in Winnipeg, Canada [7]. This data was collected from 76 vehicles with users across a range of population demographics. The data included timestamped GPS-based logs and speed data for each journey made within the data collection period. Parking events were labelled with a location category such as 'home' or 'street'. Activity profiles were constructed from this data by denoting the vehicle as 'in use' based on significant GPS movement relative to the previous recording and nonzero speed at each recorded timestamp. Where the GPS data and speed data showed different results, it was assumed that sensor error had occurred, and the data was omitted from the profile.

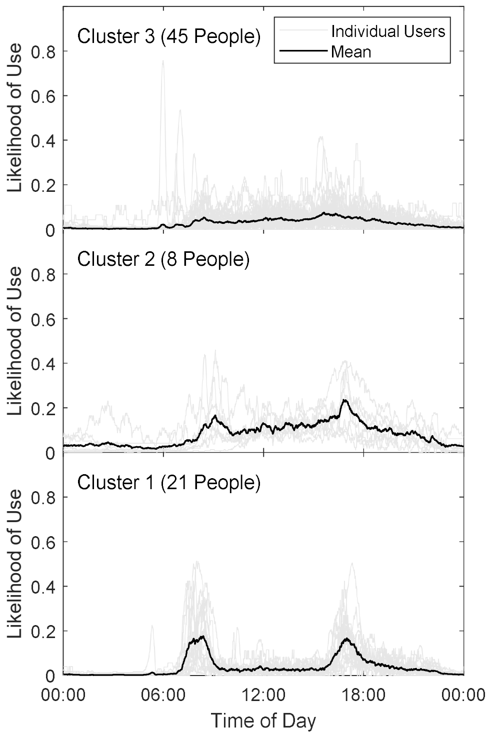

Fig. 1 Clustered Domestic Driving Activity Profiles, Winnipeg Dataset

In order to validate the choice of domestic use profiles constructed in the Agent-Based Model as discussed in Section V, clustering analysis was performed on this dataset. A k-means method was used on the calculated activity profiles to examine the most common activity demographics. After several trials with varying numbers of target clusters, it was found that three major clusters emerged from the analysis, as shown in Fig. 1. These profiles were matched roughly to three activity types: low general use typical of a vehicle not used for a work commute cycle; semi-structured use typical of users with a varying schedule; and highly structured use typical of a work commute cycle.

B. *Simulated Data*

Data for both the baseline driver archetypes and a number of varied simulated users were generated using an Agent-Based Model as detailed in Section V.

C. *Personal Device/Bluetooth Beacon Data*

Alternative vehicle use data was collected for a single participant via the driver's personal smart device. A log of driving activity was made available by placing a Kontakt Bluetooth iBeacon [8] in the participant's vehicle. The beacon emits a uniquely identifiable signal at an interval of several signals per second, which can then be detected by any device listening for Bluetooth iBeacon signals. It was assumed that the user was driving the vehicle whenever their device detected the vehicle's beacon. This data was collected via a third-party app [9] with functionality for both GPS-based geofencing and local iBeacon logging. Prior to data collection, the participant registered the Bluetooth beacon and several geofenced areas such as their home and work locations as monitored zones in the application. The data output from this application was formatted as a series of detection events as the user moved between monitored areas, including the detected location name and an entry/exit time for each event. Activity profiles were constructed from the 'in-use' data from the beacon logging: allowing this data to be analyzed alongside the vehicle-GPS data despite the difference in format.

It should be noted that the personal device/beacon setup collects data from individual drivers rather than from the vehicle itself. This means that data must be collected from all drivers in order to receive a comprehensive log of the vehicle's true use profile. The data used in this study was collected from a vehicle with a single active driver, however for multi-user vehicles, this approach introduces potential uncertainty in the usage data.

V. Agent-Based Model Archetype Generation

An agent-based model was developed to simulate archetypal driving behaviour profiles. Using this model, a number of users or 'agents' were generated, each with their own set of behavioural rules.

This model aimed to simulate driving patterns for any number of individual agents over a defined time period. In order to simplify this process, the full physical location of users was not considered. Users were instead assigned a set of location categories such as 'home' or 'work'. Journeys between these locations were represented by their duration, rather than by a physical distance. This allows realistic vehicle use patterns to be simulated without the need to calculate large volumes of co-ordinate location data.

Agent behaviours were modelled from a combination of core and additional behaviours. Core behaviours of each of the simulated domestic user archetypes are detailed in TABLE I. The archetypes in this table were designed to cover the majority of typical domestic driving styles and were validated using the clustering analysis on real driving data as discussed in Section IV. Four major archetypes were constructed, with some sub-types included as minor archetypes as detailed in the table. Each of the simulated agents were generated from one of the described archetypes. Factors such as the typical time a full-time agent leaves home for work were generated to be unique to each agent.

The agent-based simulation comprised the following main steps:

- Define the number of agents from each archetype and the time period to be simulated.
- Generate individual agent properties from archetype templates, assign each agent to a simulated vehicle.
- At the start of each simulated day, generate a set of intended journeys for each of the agents. These include core journeys based on agent archetype and additional journeys randomly drawn from a distribution that favours the afternoon-evening period.
- When the simulation reaches the start of an intended journey, assign this journey to the correct vehicle and increment journey duration and vehicle fuel level until the journey has ended. Update location category of vehicle and user.

The simulation outputs included the in-use status and location category of each of the simulated vehicles over the defined time period. The in-use status was used to calculate statistical vehicle activity profiles as defined in Section IV. An example of the use profile for a 'full-time work' agent generated from the model is presented in Fig. 2. It can be seen that the generated profile takes a similar shape to the full-time workers identified in the Winnipeg dataset in Fig. 1, with distinct peaks in activity in the morning and evening.

TABLE I. Domestic User Driving Pattern Archetypes

Major Class Archetype	Main Journeys	Main Locations	Minor Class Sub-types
Full-time worker	• 2 journeys • All Weekdays • Morning/Afternoon • Regular timing	Work Home Oher	• Base • Additional AM • Additional PM • Additional Lunch
Non-regular work	• 2 journeys • Some weekdays • Less regular times	Work Home Other	• Base
Non-worker	• No regular journeys	Home Other	• Base • Flat profile • Any time
Shift worker	• 2 journeys • All Weekdays • Offset Full-time pattern	Work Home Other	• Base

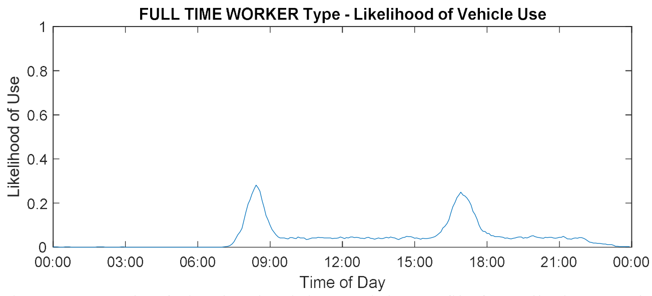

Fig. 2 Example of Simulated Driving Activity Profile for Full-Time Work

While the Agent-Based Model was used in this study to generate the baseline archetypal profiles for comparison to real user data, the simulation techniques can also be used to generate large sets of user driving patterns for other applications, including the analysis of the level of activity at different times of day over a large group of diverse users.

VI. User Archetype Classification

Once archetypal profiles were established, new users could be matched to a corresponding archetype. The advantages of this approach are as follows:

- Allows assumptions based on archetypes to be applied to new users (for example, likely location category of vehicle at given times if location data is not available, or allowing feedback into how a user's driving pattern affects their suitability for driving an electric vehicle of a given maximum range).
- Allows a less noisy activity profile to be estimated for users from a short sample of vehicle use data.
- Potential for representation/storage of user as a distortion/combination of archetypes, rather than explicitly storing and transmitting historic user location data.

In the following sections, a range of classification methods were tested on both simulated and real driving pattern data. The simulated data was used for a measure of correct classification rate, as each simulated user was defined as a specific archetype. For the real dataset, user data was anonymised, meaning that the true demographic of each user was not known. This data was therefore used to test the number of days taken to settle on a consistent classification.

A. Dynamic Time Warping

Several methods were considered for the classification process. One factor of particular importance was the ability to compare the shape of a driving profile without relying on direct comparison of each timeslot of the day. This allows a worker making their morning commute at 07:30 or a worker commuting at 09:00 to both be sorted into the 'Full-time Work' archetype, despite beginning their journey cycles at different times of the day.

This issue was addressed with the use of a Dynamic Time Warping (DTW) algorithm. This technique allows two vectors to be stretched along a single axis to find the minimum distance between points of one vector to the other. In effect, this algorithm compares the shape of two vectors and outputs the calculated distance between them at their optimum match. Further detail on the exact algorithm used in this study can be found in the MATLAB documentation [10]. The calculated distance between vectors could then be used as some measure of the 'fit' to different archetypes, where a larger distance between user data and an archetype suggests a poorer fit to this archetype. In this study, activity profiles were all scaled to have a maximum value of 1 to allow comparison of patterns without bias from variations in overall usage rates. Fig. 3 shows results of dynamic time warping on an example user from the Winnipeg dataset, as compared to the minor archetypes described in TABLE I. The sections of this figure are: Fig. 3a) unaltered user profile plotted against the unaltered matched archetype, Fig. 3b) post-DTW warped profile and archetype with optimum match levels, and Fig. 3c) a bar chart of the relative level of 'fit' of the user to each of the established archetypes. This level of fit is calculated as the inverse of the DTW distance, so that a higher value denotes a closer match.

The time-warping comparison to archetype was tested on data from the simulated, GPS/speed and personal/beacon datasets. The estimation of class was taken as the archetype with the smallest calculated distance. TABLE II. presents results from this analysis. For the simulated and personal/beacon data, where the true demographics of the vehicle drivers were known, Minor Class was considered correct if the estimated archetype was an exact match to the true demographic, while Major Class was considered correct if the estimated archetype was any sub-type of the same major archetype as the true demographic. The results show that the time-warping method is capable of correctly matching a majority of profiles to their correct archetype, however the number of days of data needed to achieve this majority was considered to be too high for practical application: it is unreasonable in most applications to expect a system to gather over one year of data before producing accurate results.

TABLE II. Dynamic Time Warping Classification Results

Dataset	Correct Minor Class Rate	Correct Major Class Rate	Mean No. Days to 70% Rate	Mean No. Days to Stabilize
Simulated	71%	84%	395	398
Vehicle GPS/Speed	-	-	-	92
*Personal GPS/ Beacon	-	-	68	76

*Single participant

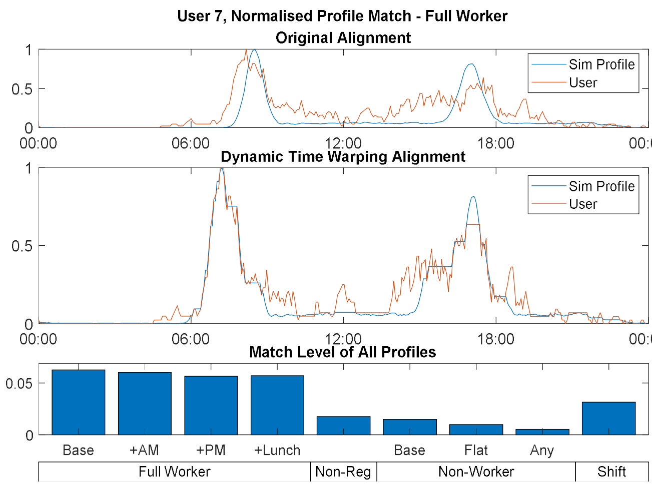

Fig. 3 Dynamic Time Warping Match of Real User Data to Archetype

It should be noted that the personal beacon dataset tested here was collected from a single participant, meaning that, while this user was correctly classified, a larger scale study would be needed for a meaningful overall classification rate.

In order to improve the classification performance with fewer days of data, several classification models were tested with a larger range of inputs than the DTW results alone.

B. Classification Model

While the Dynamic Time Warping output alone could be used to classify new activity profiles, trained classification models were tested for improvement on the number of days of data needed to reach accurate results. The classification models tested were selected based on the ability to perform multi-class classification problems and output a statistical confidence level on each of the possible output classes. Some common classification models used for multi-class applications are:

- Naïve Bayes (NB) - generally favored in cases that require a simple structure and quick implementation [11]. The major assumption applied to this model type is that each of the input features can be considered to be independent given the outcome class. However, even when this assumption is not true, Naïve Bayes models have been observed to perform well against other methods [12].
- K-Nearest Neighbors (kNN) - a non-parametric method based on the most similar available examples from the training set [13].
- Decision Tree/Random Forest (RF) - a non-parametric method, decision trees are able to represent highly nonlinear relationships without manual specification of the expected complexity of the model [14].

For each of the model types tested, simulated user data was split into two phases: training data and test data. Data from 90 users simulated over a 10-year period, with an even number of users in each sub-archetype as defined in TABLE I. 50% of the users were assigned for the training phase, where the models were constructed to best fit the training data. The models were then tested on the remaining data and the results compared to the true archetypes for each test user.

The input features tested were:

- Dynamic Time Warping distance to each archetype activity profile
- Percentage of time spent at 'home' location
- Percentage of time spent in use
- Mean no daily journeys
- Mean journey duration
- Mean overall likelihood of vehicle use.

While models trained on a full set of these features did improve the time needed to reach a given level of accuracy relative to DTW estimation alone, the features were examined to determine their contribution to the classifier performance.

C. Feature Selection

As a way to assess the relative importance of each of the input features to the overall performance of the classification models, a series of tests were performed with classifiers trained on different sets of features. The percentage change in the classifier performance when each feature was omitted from the full set is presented in TABLE III. From these results, it can be deduced that the dynamic time warping match contributes significantly to the classifier performance. The differences made by the other features are all significantly smaller, suggesting some amount of redundancy in the full set of additional features that could overcomplicate the model. Journey duration and overall mean likelihood of movement did not contribute to correct classification. The proposed optimal set of input attributes were:

- DTW distance of each sub-profile
- Percentage of time observed at 'Home' location
- Mean daily number of journeys.

The results of this optimized feature set tested on simulated users of known archetype are shown in Fig. 4, Fig. 5 and TABLE IV. While all of the trained classifiers can reach a correct classification rate with significantly fewer days of data than the time-warping activity profile comparison alone, of the model types tested there is not a clear best performer. The Random Forest model was able to achieve the highest correct classification rates by the end of the training period, but its performance was unstable over time, as indicated by the greater fluctuation than the other models seen in Fig. 4 and Fig. 5. The Naïve Bayes model achieved the highest correct rate within the first ten days of training data: in applications that require quick characterization of user patterns, this may take priority over the slight decrease in longer-term performance.

Whichever classifier is used in application, the significantly faster matching process and high correct match rates suggest that the classification approach taken in this work is feasible as a method to characterize vehicle use patterns without explicit storage of more sensitive user location data.

TABLE III. FEATURE SELECTION RESULTS (K NEAREST NEIGHBORS)

Feature Omitted	Correct Minor Class Rate Change	Correct Major Class Rate Change
Time-Warping Distance	-59%	-20%
% Time 'Home'	0%	0%
% Time 'In-Use'	0%	0%
Mean No. Journeys	-3%	-3%
Mean Duration	+17%	+3%
Mean Likelihood of Vehicle Use	+10%	0%

TABLE IV. CLASSIFIER RESULTS ON SIMULATED USERS, 2610 DAY TRIAL

Classifier	Correct Minor Class Rate	Correct Major Class Rate	No. Days to 80% Major Rate	Mean No. Days to Stabilize
DTW	71%	84%	349	398
KNN	80%	93%	2	332
NB	76%	96%	2	444
RF	82%	98%	3	2170

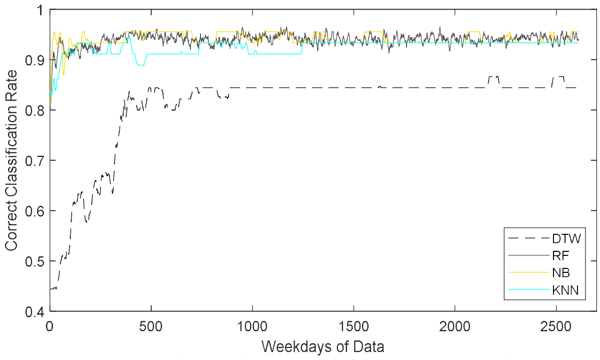

Fig. 4 Comparison of Classifiers - Rate of Correct Major Class

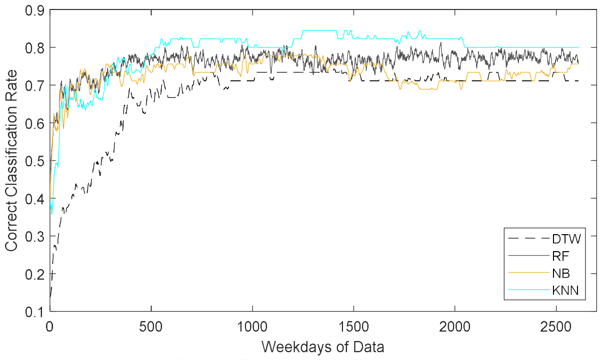

Fig. 5 Comparison of Classifiers - Rate of Correct Minor Class

VII. Conclusions

Given the increasing availability of user location data sources, vehicle usage data requires analytical approaches that are sensitive to the uncertainties involved in data collection and respond appropriately to data privacy concerns. Vehicle driving data can be collected from a range of data sources, but for wide-scale application it may not be feasible to rely on consistent data from all sources. Where wide-scale data is collected from diverse sources, this indicates a need to process data into a format that is more general. User location data also presents significant issues around user privacy and the need to securely store and transmit any personally identifiable data.

A system was proposed to allow vehicle asset use patterns to be characterised in a way that does not explicitly store large volumes of location data and allows for analysis of driving patterns relevant to a range of vehicle behaviour analysis or prediction applications. User driving data from diverse sources was converted to a statistical activity profile of vehicle usage over a 24h period.

An Agent-Based Model was used to generate a set of driving archetypes, which were validated against actual domestic vehicle driving data and were used as a basis of comparison for new driving data profiles.

Dynamic Time Warping was used to quantify the match between a new user's profile and the established driving archetypes, but this approach alone took unrealistically large volumes of data to achieve the desired level of matching accuracy. Other input features were tested in trained multi-class classification models (K Nearest Neighbours, Naïve Bayes, Random Forest). This was found to improve the number of days of data needed to make a match for most users. The set of features needed to best classify domestic driving archetypes was the time-warped match to archetypes, percentage of time in 'home' location and mean daily number of journeys, with the Random Forest model giving a correct classification rate of 82% in a test on simulated data.

The archetyping methods proposed in this paper have potential for use across a range of application fields, including to inform the prediction of vehicle dwell and journey times. Further testing with large datasets of real driving data would better inform the method and ensure a comprehensive set of driving archetypes were developed.

Acknowledgment

In addition to thanking Innovate UK and A.T. Kearney for their sponsorship of this work, the authors would like to thank the providers of the open-source driving dataset in [7].

References

[1] J. Guinn, 'Fleet Management Software Buyer's Guide', 04-Jun-2018. [Online]. Available: https://www.softwareadvice.com/uk/fleet-management/#buyers-guide. [Accessed: 08-Jun-2018]

[2] Y. Kuang, Y. Chen, M. Hu, and D. Yang, 'Influence analysis of driver behavior and building category on economic performance of electric vehicle to grid and building integration', *Applied Energy*, vol. 207, pp. 427–437, Dec. 2017 [Online]. Available: http://www.sciencedirect.com/science/article/pii/S030626191730870X

[3] R. Kirk, 'Cars of the future: the Internet of Things in the automotive industry', *Network Security*, p. 3, 2015.

[4] 'Personal Data – GDPR EU.org'. [Online]. Available: https://www.gdpreu.org/the-regulation/key-concepts/personal-data/. [Accessed: 27-Sep-2018]

[5] Y.-A. de Montjoye, C. A. Hidalgo, M. Verleysen, and V. D. Blondel, 'Unique in the Crowd: The privacy bounds of human mobility', *Scientific Reports*, vol. 3, p. 1376, Mar. 2013 [Online]. Available: https://www.nature.com/articles/srep01376. [Accessed: 10-Sep-2018]

[6] S. Wachter, 'Normative challenges of identification in the Internet of Things: Privacy, profiling, discrimination, and the GDPR', p. 14, 2018.

[7] A. Ashtari, Eric Bibeau, and S. Shahidinejad, 'WPG01 driving cycle'. 2010 [Online]. Available: http://dx.doi.org/10.5203/ds_bib_2. [Accessed: 07-Sep-2018]

[8] Kontakt, 'Beacon', 2016. [Online]. Available: https://store.kontakt.io/our-products/27-beacon.html. [Accessed: 25-Jan-2017]

[9] Tion, 'Geofency for iPhone - Mobile Time Tracking', 2013. [Online]. Available: http://www.geofency.com. [Accessed: 15-Feb-2017]

[10] 'Distance between signals using dynamic time warping - MATLAB dtw - MathWorks United Kingdom'. [Online]. Available: https://uk.mathworks.com/help/signal/ref/dtw.html. [Accessed: 21-Sep-2018]

[11] E. Frank, L. Trigg, G. Holmes, and I. H. Witten, 'Technical note: Naive Bayes for regression', *Machine Learning*, vol. 41, no. 1, pp. 5–25, 2000 [Online]. Available: http://www.springerlink.com/index/h6nt787323166532.pdf. [Accessed: 21-Jun-2017]

[12] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer Science & Business Media, 2001 [Online]. Available: https://web.stanford.edu/~hastie/ElemStatLearn/printings/ESLII_print12.pdf. [Accessed: 26-Jun-2018]

[13] Scikit-learn Developers, 'Nearest Neighbors', 2016. [Online]. Available: http://scikit-learn.org/stable/modules/neighbors.html. [Accessed: 15-Jun-2017]

[14] Analytics Vidhya Content Team, 'A Complete Tutorial on Tree Based Modeling from Scratch', *Analytics Vidhya*, 12-Apr-2016. [Online]. Available: https://www.analyticsvidhya.com/blog/2016/04/complete-tutorial-tree-based-modeling-scratch-in-python/#one. [Accessed: 14-Jun-2017]