Vehicle Classification from Low Frequency GPS Data

Matteo Simoncini, Francesco Sambo, Leonardo Taccari, Luca Bravi, Samuele Salti and Alessandro Lori
Fleetmatics Research
Viale Mazzini 40, Florence, Italy
Email: alessandro.lori@fleetmatics.com

***Abstract*—Inferring the type of vehicles on a road is a fundamental task within several applications. Some recent works have exploited Global Positioning System (GPS) devices and used classification of GPS traces to tackle the problem. Existing approaches based on GPS data make use of GPS trajectories sampled at high frequency (about 1 sample per second), but GPS trackers currently installed on public and commercial fleets acquire GPS positions at lower frequency (about 1 sample per minute).**

In this paper, we target the more challenging scenario of low frequency GPS data, which has not been tackled yet in the literature, and explore how this kind of data can be used to effectively categorise vehicles into light-duty and heavy-duty. We define several distance-, speed-, and acceleration-based features, inspired by the literature on related problems like travel mode detection, and add novel features based on road type. Features are aggregated over a GPS track with several aggregation functions. We identify the most effective combinations of features and aggregation functions with a data-driven approach, by applying Recursive Feature Elimination in a cross validation framework. Furthermore, we combine predictions of all tracks of a vehicle to boost classification performance. Experimental results on a large dataset show that the selected features are indeed effective and that the high and low frequency GPS scenarios greatly differ in terms of relevant features.

I. Introduction

Inferring the type of vehicles in a road network, a problem typically referred to in the literature as *vehicle classification*, is a fundamental task in several applications, such as, *e.g.*, surveillance systems [1], traffic management [2], emission control and estimation of highway lifespan [3]. The Federal HighWay Administration (FHWA) of the United States proposed a 13-category vehicle taxonomy, based on the vehicle weight, length, axles number and axles distances [4]. Even if the rules have been revised over the years by companies and agencies [5], the FHWA 13 vehicle categories are still used as a classification target.

Methods to address the vehicle classification problem employ both hardware and software solutions in different contexts. When physical components can be installed along a road, hardware-based techniques using fixed-location sensorscan be adopted [3], [6]. Such approaches can provide the full 13-class classification in exchange for a high installation cost. On the other hand, most software-based techniques perform classification from images or videos, obtained, for instance, from surveillance cameras. The classification targets can vary, also depending on the resolution of the cameras: *e.g.*, vans, taxis and passenger cars are considered in [1], while sedans, pickups and vans in [7].

In the last decade, the great diffusion of GPS (Global Positioning System) devices is generating a growing interest in the application of data mining techniques to the huge amount of spatio-temporal data produced by such devices. GPS data are typically produced by either general-purpose mobile devices (*e.g.*, smartphones) or dedicated GPS tracker devices, usually installed on commercial or public transport vehicles (*e.g.*, delivery fleets, taxis, ambulances) [8]. In the former case, GPS signals are used for navigation or geo-localisation purposes, hence with high sampling rates (of the order of one GPS sample per second). In the latter case, GPS signals are typically used for remote vehicle tracking or anti-theft systems and lower frequency sampling (of the order of one sample per minute) is sufficient. The use of low frequency GPS data allows for the reduction of operational costs due to bandwidth, storage space and computational power and is therefore very common in industrial applications and commercial fleet management solutions [8]. Clearly, the technical and economical advantages come at the cost of accuracy: lower frequency sampling means that information on instantaneous speeds and accelerations are scarce or not available at all and that it is harder to infer the true path of a vehicle between two reported positions.

The ubiquity of GPS devices motivates the use of methods that leverage GPS data to perform vehicle classification. Relying exclusively on GPS-based features makes a complete 13-category classification very challenging, due to the difficulty to have a clear measure of axles number and distances. However, as reckoned in [3], in many real traffic applications, such as travel pattern or quality of traffic flow estimation, it is often sufficient to be able to distinguish between two or three classes of vehicles.

To the best of our knowledge, the only work that explores vehicle classification from GPS data [3] considers a two-class classification problem, distinguishing between passenger cars and delivery trucks. The reported results are obtained from a small size dataset comprising 52 tracks of passenger cars and 84 tracks of trucks. GPS data used in the paper has a sample rate of 3 seconds, which is relatively high. The authors conclude that speed-related features greatly depend on traffic conditions, whereas acceleration- and deceleration-based features have a more consistent predictive power.

Closely related to vehicle classification is the problem of travel mode detection [9]–[12], albeit with the substantial difference that the detection of travel modes such as *walk*, *bus*, *train* and *car* allows for the use of highly discriminating features (such as speed, number of heading changes, number of stops, distance travelled) that may not be equally effective in discriminating more finely-grained vehicle classes. In spite of this difference, some features used for travel mode detection could turn out useful also for the vehicle classification problem. For example, in [9], average speed, 95th percentile speed, average absolute acceleration and travel distance are used; average and maximum speed and accelerations and total distance travelled were proposed in [10]; mean speed, top three speeds of the segment and speed standard deviation were shown to be effective in [11].

As far as low frequency data are concerned, they are used for travel mode detection only by [12] (every 60 seconds). The authors consider a variable-size moving window on both acceleration and speed of a GPS track to predict the travel mode. Low frequency GPS data have also been used for the problem of travel time estimation [13], [14].

In this paper, we address the unexplored problem of vehicle classification from low frequency data, such as the data provided by GPS devices installed in commercial fleets for vehicle-tracking purposes. Rather than focusing on classifying passenger cars and trucks, as in [3], we aim at performing binary vehicle classification over a more heterogeneous range of road vehicles, distinguishing between smaller, light-duty vehicles (*i.e.*, cars, SUVs, vans and light duty pickups, that correspond to classes 2-3 of the FHWA scheme [4]) and larger size vehicles (*i.e.*, heavy duty pickups, small trucks, trucks and big trucks, classes 5-12 of the FHWA scheme).

In order to define a baseline for our approach, we adapt the existing method [3], proposed for high frequency data, to our scenario. Together with the acceleration-based features identified in [3] as highly predictive, we consider also the speed and distance-related features exploited for travel mode detection, as they might become more relevant in the low frequency scenario.

The main contributions of this paper are the following:

- we propose a comprehensive set of features based on speed, distance and acceleration, inspired by state-of-the-art algorithms on travel mode detection, and information on the type of travelled road;
- we exploit a purely data-driven approach to assess the relative importance and predictive power of each feature, with a recursive feature elimination procedure applied to a large vehicle dataset (more than 100 000 GPS tracks of about 2000 vehicles), obtaining better performance than the current state of the art;
- by analysing the ranking of the selected features, we provide insight into the kind of features that are relevant when classifying vehicles in a low frequency sampling scenario;
- we show how classification performance can be significantly improved by aggregating classification scores across multiple tracks of the same vehicle.

The remainder of this paper is structured as follows. Section II presents the methodology, in particular Section II-A provides a high-level description of GPS data, Section II-B presents in more details the method described in [3] and discusses how it can be adapted to low frequency data, in order to provide a baseline for our approach, and Section II-C introduces all the features we evaluate and the feature selection procedure. Section III presents the dataset used for the evaluation and discusses the experimental results. Finally, Section IV reports conclusions and future directions.

II. Methods

In this Section, we describe the structure of GPS data we used and how the solution described in [3] can be adapted to a low frequency scenario. Afterwards, we introduce the approach proposed in this paper, based on the application of a feature selection algorithm to a broad set of features which may have predictive power for vehicle class prediction from low frequency GPS data.

A. GPS data

A *GPS track* is a sequence of GPS *samples* (or points) $\{\mathbf{P}_i\}_{i=1}^n = \{\mathbf{P}_1, \ldots, \mathbf{P}_n\}$, where $\mathbf{P}_1$ is the first point, obtained immediately after the engine is turned on, and $\mathbf{P}_n$ is the last point, obtained just before the engine is turned off. Each GPS point $\mathbf{P}_i$ contains latitude and longitude (*i.e.*, position $\mathbf{p}_i$), odometer distance d_i, timestamp t_i and instantaneous speed v_i. For future reference, we also define $\{\mathbf{p}_i\}_{i=1}^n$, $\{d_i\}_{i=1}^n$, $\{t_i\}_{i=1}^n$, $\{v_i\}_{i=1}^n$ as the sequence of positions, distances, times and speeds of a GPS track, respectively. It is worth pointing out that we cannot assume uniform sampling rates in a track or across tracks, as data collected by heterogeneous GPS devices may have different sampling rates and the sampling rate can vary in the same device according to vehicle speed or the occurrence of asynchronous triggers, like *e.g.* harsh driving events. Fig. 1 reports an example of a GPS track.

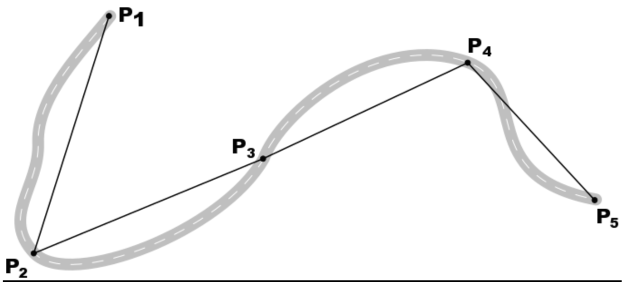
Starting from the raw data sequences, we introduce the derived measurements $\{\tilde{v}_i\}_{i=1}^n$, $\{a_i\}_{i=1}^n$ and $\{\tilde{a}_i\}_{i=1}^n$ as the sequence of finite differences of distance and of instantaneous speeds, and the second order finite differences of speed over time, *i.e.*,

$$\tilde{v}_i = \frac{d_i - d_{i-1}}{t_i - t_{i-1}} \ , \tag{1}$$

$$a_i = \frac{v_i - v_{i-1}}{t_i - t_{i-1}} \ , \tag{2}$$

$$\tilde{a}_i = \frac{\tilde{v}_i - \tilde{v}_{i-1}}{t_i - t_{i-1}} \ . \tag{3}$$

We consider both instantaneous speeds and speeds based on finite differences, computed as distance over time, as they convey different information when applied to low frequency data: instantaneous reads every 60 or more seconds can be very noisy, but may capture short parts of the track with high discriminative information, *e.g.*, high speed events, whereas



	Time	Position	Speed	Distance
$\mathbf{P}_1$	0.0 s	(42.13222, -72.55265)	0.0 m/s	0.0 m
$\mathbf{P}_2$	120.0 s	(42.13737, -72.54227)	6.1 m/s	677.99 m
$\mathbf{P}_3$	221.0 s	(42.12700, -72.54660)	16.3 m/s	1,830.12 m
$\mathbf{P}_4$	311.0 s	(42.11551, -72.55024)	17.1 m/s	2789.96 m
$\mathbf{P}_5$	402.0 s	(42.10739, -72.54664)	0.0 m/s	3792.06 m

Fig. 1. Example of a GPS track composed by five GPS samples. The sampling frequency varies depending on vehicle speed and it is affected by hardware delay, thus being not constant during the sequence. Note that the reported speed is instantaneous, so it may be affected by traffic or street conditions.

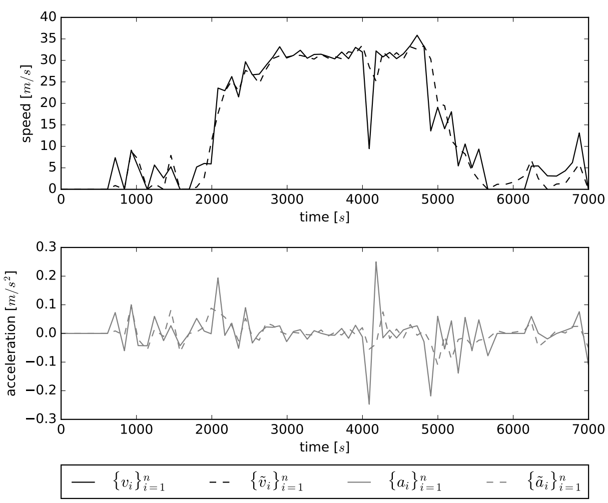


Fig. 2. Example of speed, finite difference speed, acceleration and finite difference acceleration sequences.

speed computed as the ratio of distance over time in a longer interval provides a more reliable and smooth estimate of the trend of speed along the track. Similarly, since the instantaneous acceleration is not available in our raw data, we compute it as both finite differences of instant speed and second order finite differences of distance over time. An example of the two types of speed and acceleration is given in Figure 2. For conciseness, we name $\tilde{v}$, a and $\tilde{a}$ as *interval speed*, *acceleration* and *interval acceleration*, respectively, in what follows. After this preprocessing, every GPS point $\mathbf{P}_i$ is associated to a set of values $(\mathbf{p}_i, t_i, d_i, v_i, \tilde{v}_i, a_i, \tilde{a}_i)$.

Acceleration and deceleration exhibit different patterns in the different vehicle classes, as mentioned also in [3]. However, parts of the track where the vehicle is idling[1], *e.g.* in a traffic jam, can be less useful for recognising its class. Therefore, we associate to a GPS track six more sets, $\{a_{>0}\}$, $\{a_{<0}\}$, $\{\tilde{a}_{>0}\}$, $\{\tilde{a}_{<0}\}$, $\{v_{>0}\}$ and $\{\tilde{v}_{>0}\}$, that we name *positive acceleration*, *deceleration*, *positive interval acceleration*, *interval deceleration*, *positive speed* and *positive interval speed*, respectively, defined as

$$\{a_{>0}\} = \{a_i | a_i > 0, i = 1, \ldots, n\} \tag{4}$$
$$\{a_{<0}\} = \{-a_i | a_i < 0, i = 1, \ldots, n\} \tag{5}$$
$$\{\tilde{a}_{>0}\} = \{\tilde{a}_i | \tilde{a}_i > 0, i = 1, \ldots, n\} \tag{6}$$
$$\{\tilde{a}_{<0}\} = \{-\tilde{a}_i | \tilde{a}_i < 0, i = 1, \ldots, n\} \tag{7}$$
$$\{v_{>0}\} = \{v_i | v_i > 0, i = 1, \ldots, n\} \tag{8}$$
$$\{\tilde{v}_{>0}\} = \{\tilde{v}_i | \tilde{v}_i > 0, i = 1, \ldots, n\}. \tag{9}$$

B. Baseline model

To define a baseline model for our analysis, we adapt the vehicle classification approach described in [3] to our low frequency GPS data scenario. In the original proposal, the authors first split the acceleration sequence retrieved from GPS tracks in the set of (strictly positive) accelerations and decelerations. Then, to obtain features at the track level from a sequence of GPS samples, they separately aggregate the values from the two sets with two measures:

- standard deviation of the values across the track
- fraction of values greater than 1 m/s^2.

The standard deviation of the acceleration can be computed also for low frequency data. However, note that acceleration is estimated over a much larger time interval: if we consider, for instance, a 60 seconds sampling rate, an acceleration larger than 1 m/s^2 could only be obtained observing a speed difference between two consecutive GPS samples of more than 60 m/s, *i.e.* 216 km/h, which is extremely unlikely to occur in real data. Thus, to adapt the approach in [3] to low frequency data, we consider a variable threshold T to define the informative part of the distribution of accelerations and decelerations, and we tune it via cross-validation from the training set. We assessed both the a_i and $\tilde{a}_i$ definitions of the acceleration and found that the second order version leads to slightly better classification performance, overall. Classification is then performed, as in [3], using a Support Vector Machine (SVM) [15] with quadratic kernel.

C. Feature selection approach

As discussed in the introduction, the acceleration-based features used in [3] may not be the only predictive cues with low frequency data. Speed and distance-related features, usually exploited for travel mode detection, might become more relevant. Given the differences between our task and travel mode detection, it is difficult to make any *a priori* assumption on the importance of each feature and on what aggregation function to use in our scenario. For this reason, we decided to consider a large number of features and aggregation functions and to automatically filter the redundant and not-discriminative features using a feature selection algorithm. We consider a set of 10 feature sequences for each GPS track, reported in Tables I, and aggregate them with 13 functions,

[1] We define *idling* as engine on without the vehicle moving.

TABLE I
LIST OF FEATURE SEQUENCES AND SETS OBTAINED FROM A GPS TRACK

Variable	Description
$\{v_i\}_{i=1}^n$	speed sequence
$\{\tilde{v}_i\}_{i=1}^n$	interval speed sequence
$\{a_i\}_{i=1}^n$	acceleration sequence
$\{\tilde{a}_i\}_{i=1}^n$	interval acceleration sequence
$\{a_{>0}\}$	positive acceleration set
$\{a_{<0}\}$	deceleration set
$\{\tilde{a}_{>0}\}$	positive interval acceleration set
$\{\tilde{a}_{<0}\}$	interval deceleration set
$\{v_{>0}\}$	positive speed set
$\{\tilde{v}_{>0}\}$	positive interval speed set

TABLE II
LIST OF AGGREGATION FUNCTIONS

Function	Description
$mean$	mean value of the sequence
std	standard deviation of the sequence
$median$	median value of the sequence
mad	median absolute deviation of the sequence
iqr	interquartile range of the sequence
$75th$	75th percentile of the values in the sequence
$90th$	90th percentile of the values in the sequence
$95th$	95th percentile of the values in the sequence
$hist1$	fraction of samples in the first bin of the 6-bin histogram
$hist2$	fraction of samples in the first 2 bins of the 6-bin histogram
$hist3$	fraction of samples in the first 3 bins of the 6-bin histogram
$hist4$	fraction of samples in the first 4 bins of the 6-bin histogram
$hist5$	fraction of samples in the first 5 bins of the 6-bin histogram

reported in II. For example, we consider the sequence of speeds $\{v_i\}_{i=1}^n$ of a track and aggregate it with the $mean$ function, obtaining the mean speed of the track as a feature for the whole track. This results in a total of 130 possible features for each track. To define the 6 histogram bins for the aggregation functions $hist1, \ldots, hist5$ we consider all the values of all tracks in the training set for a feature sequence, *e.g.* all values of speed in the training set, we take the interval between the 5th and 95th percentiles and divide it into 6 bins. $hist6$ is not used as feature as it always includes all samples of the sequence.

We also include in the set of features the total distance covered by the vehicles in the track, inspired by the fact that in general heavy duty vehicles are used for longer trips compared to smaller vehicles.

Finally, we also consider a set of domain-related features, *i.e.*, the category of the road on which each GPS sample is acquired, due to the fact that speed and acceleration of different vehicles may vary significantly according to the road on which they are travelling. In particular, for each track, we compute the fraction of GPS samples in the sequence falling in each of the considered road type, which are *motorway*, *highway*, *trunk road*, *country road*, *city road* and *residential road*, for a total of six additional features. Hence, the final number of considered features for each track is 137.

Feature selection is a well-known problem in supervised learning, where the aim is to reduce the dimensionality of the feature space in order to overcome the risk of overfitting, *i.e.*, of learning a model too fit to the training data and unable to generalise to unseen data. For our analyses, we chose the Recursive Feature Elimination (RFE, [16]) algorithm for feature selection: the algorithm starts from the entire set of features and tests the impact of the removal of one feature at a time on classification performance. Then, the feature whose removal leads to the smallest decrease (or the largest increase) in classification performance is removed from the set of used features. The process is then iterated until no feature remains, resulting in a ranking of features in increasing order of importance for the classification task. In order to limit computational time, the possibility of removing more than one feature at a time in each RFE iteration is suggested in [16], at the expense of a minor decrease in precision of the feature ranking.

As a measure of classification performance, we chose the widely adopted Area Under the ROC, or Receiver Operation Characteristic, curve (AUC for brevity). Such metric considers the curve of variation of *false positive rate* (*fpr*) *vs. true positive rate* (*tpr*) at different values of the classification threshold (ROC curve), where

$$\mathit{tpr} = \frac{TP}{TP + FN}\,, \qquad \mathit{fpr} = \frac{FP}{FP + TN}\,. \tag{10}$$

The area under the ROC curve depends on both false positives and true positives: this lets it penalise models which are representative but not discriminative and, thus, makes it robust even in the case of unbalanced datasets.

The pseudocode of our RFE procedure is reported in Algorithm 1. The dataset is first split by performing k-fold cross validation, stratified on the two classes, to be able to statistically assess classification performance on k sets of independent data (line 1 of Algorithm 1). For each cross-validation train/test pair, we halve the number of features at each iteration of RFE by discarding those features whose removal yields to the lowest AUCs. Therefore, given the number of input features m, we can pre-compute the number of iterations $\overline{m}$, line 2, and the number of features to be removed at each iteration, array B, line 4. At each RFE iteration, to compute AUCs we first set aside one feature at a time from the set of remaining features, train the model on the train set and assess the corresponding AUC on the test set (line 10).

At the end of this procedure, we obtain a ranked list of features for each fold, stored in the rows of the matrix of ranks R (line 14). To compute a ranking over the whole training set, we sort features according to their median rank across the k folds (line 19). To compute the optimal number of features f^*, we consider the sets $\{AUC_{1,t}, \ldots, AUC_{k,t}\}$ of

Algorithm 1 Recursive Feature Elimination

Input: Dataset $D \in \mathcal{M}_{n\times m}(\mathbb{R})$
Output: Best feature set
1: $(\mathit{Train}_1, \mathit{Test}_1), \ldots, (\mathit{Train}_k, \mathit{Test}_k) \leftarrow \mathit{kFoldCV}(D)$
2: $\overline{m} \leftarrow \lceil \log_2 m \rceil + 1$
3: Define $R \in \mathcal{M}_{k\times m}(\mathbb{R})$, $\mathit{AUC} \in \mathcal{M}_{k\times \overline{m}}(\mathbb{R})$, $B \in \mathbb{R}^{\overline{m}}$, $R^{\mathit{med}} \in \mathbb{R}^{\overline{m}}$
4: $B \leftarrow \{m, \lceil m/2 \rceil, \left\lceil \frac{\lceil m/2 \rceil}{2} \right\rceil, \ldots, 1\}$
5: **for** $i \leftarrow 1$ **to** k **do**
6: $\quad$ *features* $\leftarrow \{1, \ldots, m\}$
7: $\quad$ $\mathit{AUC}_{i,1} \leftarrow$ AUC of the classifier trained on Train_i and tested on Test_i, with features *features*
8: $\quad$ **for** $b \leftarrow 2$ **to** $\overline{m}$ **do**
9: $\quad\quad$ **for** $j \leftarrow 1$ **to** *length* (*features*) **do**
10: $\quad\quad\quad$ Compute AUC of the classifier trained on Train_i and tested on Test_i, with features *features* $\setminus j$
11: $\quad\quad$ **end for**
12: $\quad\quad$ *features* $\leftarrow$ Best B_b features
13: $\quad\quad$ $s \leftarrow 1 + \sum_{i=2}^{b-1} B_i, \quad e \leftarrow \sum_{i=2}^{b} B_i$
14: $\quad\quad$ $R_{i,s\ldots e} \leftarrow \{$worst B_b features, ordered by AUC$\}$
15: $\quad\quad$ $\mathit{AUC}_{i,b} \leftarrow$AUC of the classifier trained on Train_i and tested on Test_i, with features *features*
16: $\quad$ **end for**
17: **end for**
18: **for** $b \leftarrow 1$ **to** $\overline{m}$ **do**
19: $\quad$ $R_b^{\mathit{med}} \leftarrow$ median rank across $R_{1,b}, \ldots, R_{k,b}$
20: **end for**
21: $\mathit{pVal} \leftarrow 0$, $t \leftarrow \overline{m}$
22: **while** $\mathit{pVal} < 0.05$ **and** $t > 1$ **do**
23: $\quad$ $\mathit{pVal} \leftarrow$ p-value of a Wilcoxon signed-rank test for significantly larger median between the AUC distributions $\{\mathit{AUC}_{1,t}, \ldots, \mathit{AUC}_{k,t}\}$ and $\{\mathit{AUC}_{1,t-1}, \ldots, \mathit{AUC}_{k,t-1}\}$
24: $\quad$ $t \leftarrow t - 1$
25: **end while**
26: $f^* \leftarrow B_{t+1}$
27: **return** f^* features with the lowest median rank R_{med}

AUCs obtained at each iteration t of the RFE procedure. We compare each pair of sets of k AUCs, obtained with feature sets of increasing size, with a Wilcoxon signed-rank test for significantly larger median [17], stopping when the test detects no significant increase in AUC, with confidence threshold 0.05 (lines 21 to 26).

As a classifier we use Support Vector Machines (SVM [15]) for their recognised effectiveness in binary classification [16]. The optimal configuration of the SVM model, namely the kernel type and its numerical parameters, is chosen via k-fold cross validation, as explained in the next section.

III. Experimental results

A. Data collection and description

To our knowledge, no public low frequency GPS datasets labelled by vehicle type exist. Our dataset was collected by Fleetmatics over two months of activity of vehicles in the USA. The GPS sampling rate of the devices varies depending on the status of the vehicle: if the speed is lower than a threshold while the engine is on, the vehicle is idling and the sampling interval is $120s$, otherwise the vehicle is moving and the sampling interval is $90s$.

To build the dataset, a subset of vehicles tracked by Fleetmatics were manually labelled by considering the maker and model reported by the owner, dividing the data as follows:

- **CAR**, compact and subcompact cars
- **SUV**, SUVs and jeeps
- **VAN**, vans and commercial vans
- **PICKUP**, pickups with engine capacity lower than 3.5 litres
- **SMALLTRUCK**, pickups with engine capacity higher than or equal to 3.5 litres, small sized trucks
- **TRUCK**, medium sized trucks
- **BIGTRUCK**, big sized trucks.

The labelling was mainly performed by visual inspection of the model of the vehicle and, thus, it depends on some judgement calls on the borders (*e.g.*, the distinction between medium and big trucks). It is also worth noting that several car companies implement cutaway versions of vans and pickups, making them more similar to small trucks than pickups or vans. Finally, small pickups are used similarly to cars or vans mainly for people or small equipment transportation while heavy duty pickups, designed for the transportation of heavy loads, are more similar in utilisation and GPS dynamics to small trucks. Therefore, we chose to put a threshold on the engine displacement, and to label pickups with engine displacement smaller than 3.5 litres as *PICKUP*, and as *SMALLTRUCK* otherwise.

The vehicles were further grouped in two higher level classes in order to perform binary classification: *CAR*, *SUV*, *VAN* and *PICKUP* were labelled as **LIGHT-DUTY**, *SMALLTRUCK*, *TRUCK* and *BIGTRUCK* were labelled as **HEAVY-DUTY**.

For the present analysis, we decided to target the Small and Medium Businesses (SMBs) segment: 1987 vehicles were sampled from the set of Fleetmatics SMB customers stratifying on vehicle type, *i.e.*, reproducing the distribution of the vehicle types in the business segment. The distribution of the vehicles across types in the sampled dataset is reported in Figure 3: as it is clear from the figure, the vast majority of vehicles are *PICKUP*s and *VAN*s.

For each vehicle, the full set of GPS samples over two months were segmented into GPS tracks by using `engine on` and `engine off` events triggered by the GPS devices. Only vehicles having at least 10 GPS tracks in the two months were considered. Furthermore, only tracks having at least 3 positive accelerations and 3 decelerations were considered, *i.e.*, $|\{a_{>0}\}| \geq 3$, $|\{a_{<0}\}| \geq 3$, $|\{\tilde{a}_{>0}\}| \geq 3$ and $|\{\tilde{a}_{<0}\}| \geq 3$. The final dataset is formed by $15\,792$ GPS tracks in the HEAVY-DUTY class and $103\,122$ GPS tracks in the LIGHT-DUTY class. We split the dataset in a training and a test set, composed of 993 and 994 vehicles and $58\,525$ and $58\,453$ tracks, respectively.

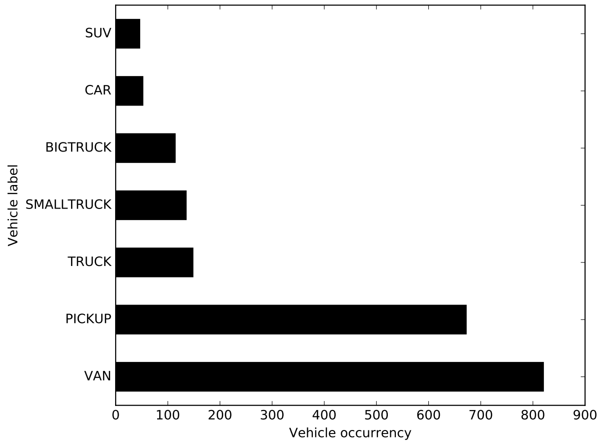


Fig. 3. Vehicle distribution in the dataset.

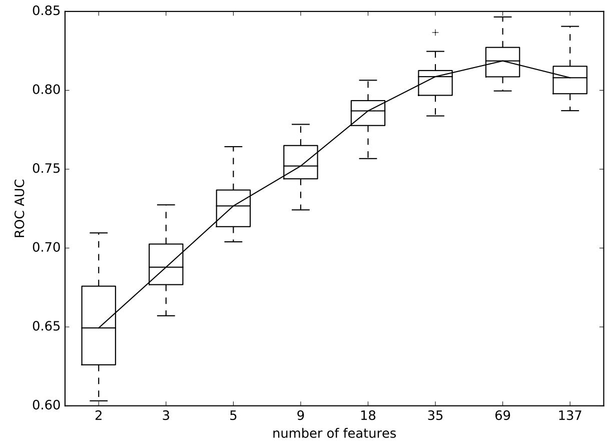


Fig. 4. Box plots over the 15 folds of the ROC AUCs obtained with Recursive Feature Elimination, for increasing number of features.

TABLE III
PERFORMANCE OF KERNEL FUNCTIONS BEFORE FEATURE SELECTION, ESTIMATED BY 5-FOLD CROSS VALIDATION ON THE TRAINING SET

Kernel	Optimal parameters	ROC AUC
Linear	$C = 1$	0.6881
Polynomial	$C = 1$ $c_0 = 1$ $\gamma = 0.01$ $d = 3$	0.7973
RBF	$\mathbf{C = 46.4159}$ $\boldsymbol{\gamma = 0.0022}$	**0.8036**

TABLE IV
OPTIMAL PARAMETERS USED IN THE EXPERIMENTS.

Method	Optimal parameters
Baseline	$C = 0.5623$ $T = 0.375$
Proposed Method	$C = 46.4159$ $\gamma = 0.0022$

Finally, in order to obtain the road type information, each (latitude, longitude) pair was processed with the commercial product *PTV xServer*[2], but similar results could be obtained with any free reverse geocoding service, *e.g.*, *OpenStreetMap*[3]. As a result, the fractions of GPS samples in the train and test sets for each road type were 21.63% for *motorway*, 2.29% for *highway*, 3.41% for *trunk road*, 14.55% for *country road*, 25.96% for *city road* and 32.16% for *residential road*. These numbers reflect the operations distribution of Fleetmatics SMB customers, with $25/30\%$ long haul trips on major roads and the remainder of traffic on country roads or city networks.

B. Results

The baseline approach based on [3] was tuned using stratified 5-fold cross validation to estimate the optimal threshold T and the SVM parameter C [15] which maximise the AUC. Optimal parameters are reported in Table IV.

To select the best SVM kernel and tune its parameters for the proposed approach, we compare three commonly used kernels (linear, polynomial and Radial Basis Function, or RBF) and assess several combinations of kernel parameters with grid search, using the entire feature set in a stratified 5-Fold cross validation. The results of kernel selection and parameter tuning are reported in Table III, with RBF as the best kernel. The kernel parameters are then kept fixed during the feature selection step. When the best subset of features are identified, stratified 5-Fold cross validation is used again to obtain the final C and kernel parameters, which are reported in Table IV. To take into account unbalanced classes, the misclassification parameter C is weighted with respect to the number of training examples in each class for both methods.

The RFE algorithm has been applied with 15-fold cross validation to increase the statistical significance of the feature ranking. The box plots of the AUCs obtained in each step of the RFE algorithm are reported in Figure 4. The one-tailed paired Wilcoxon test reported a p-value always lower than $5 \cdot 10^{-5}$, except on the comparison between 69 features and 137 features ($p = 0.997$). Thus, the optimal number of features f^* was set to 69. In Table V, the list of selected features is reported, along with their ranking. The most relevant features are the travelled distance and the fraction of GPS samples in each type of road, combined with aggregated features regarding the distribution of the positive speed values (standard deviation, 95th percentile and fraction of samples in the first bins of the histogram). A few acceleration-based features, in accordance to what was found in [3], are still relevant, although they appear to be generally less predictive than speed-based measures when dealing with low frequency data.

In Figure 5, the resulting ROC curves for the baseline and the proposed method are reported. The results show that introducing new domain-based features and using data-driven feature selection leads to a significant improvement over the set of features suggested by [3] in our scenario of low-frequency GPS data (a ROC AUC of 0.769 compared to a baseline of 0.715).

It is important to stress that domain-based features alone,

[2]http://xserver.ptvgroup.com/
[3]http://www.openstreetmap.org/

TABLE V

RESULTING SET OF SELECTED AGGREGATED FEATURES AND DOMAIN-RELATED FEATURES. THE PRESENCE OF A NUMBER INDICATES THAT THE FEATURE-FUNCTION PAIR (INT THE TOP TABLE) OR THE DOMAIN-RELETED FEATURE (IN THE BOTTOM TABLE) HAS BEEN SELECTED BY THE ALGORITHM, THE LUMINANCE AND THE NUMBER OF THE ENTRY INDICATES THE RANKING, FROM THE HIGHEST (BLACK) TO THE LOWEST (GRAY).

	$\{v_i\}_{i=1}^n$	$\{\tilde{v}_i\}_{i=1}^n$	$\{a_i\}_{i=1}^n$	$\{\tilde{a}_i\}_{i=1}^n$	$\{a_{>0}\}$	$\{a_{<0}\}$	$\{\tilde{a}_{>0}\}$	$\{\tilde{a}_{<0}\}$	$\{v_{>0}\}$	$\{\tilde{v}_{>0}\}$
$mean$	19	47	68						20	36
std	13	26			54				5	27
$median$	62	50			56		42		49	39
mad				30				66		
iqr		53					64	69	59	23
$75th$				15				58	46	35
$90th$	67	51								43
$95th$	45	29							8	16
$hist1$	28	31			18	63	61		12	9
$hist2$	32	21	55						25	14
$hist3$	22	57	33	11			48		38	40
$hist4$	24		60				34			41
$hist5$	44			65			17		37	52

Road type percentage						Total distance
motorway	*highway*	*trunk road*	*country road*	*city road*	*residential road*	
3	7	10	6	4	2	1

though highly ranked, would not be able to provide a good classification performance: a simple experiment using only road type and total travelled distance yields an AUCs of 0.65, far below what we obtain when the optimal 69 features are considered.

The results shown so far involve only vehicle classification from a single observed track. However, several tracks are actually available for each vehicle in our dataset. It is thus natural to assess the effectiveness of performing vehicle classification based on the full set of GPS data at our disposal for each vehicle, rather than on a single track. As a first step, every track in the dataset is classified by using the single-track SVMs, yielding a discrete label and the continuous, signed value of the decision function for the single-track. Then, we use as decision function for each vehicle the average of the values of the single-track decision functions obtained from its GPS tracks. The results reported in Figure 6 show that the aggregation of multiple tracks for a vehicle leads to a consistent increase in classification performance (0.89 AUC).

Finally, in order to analyse the classification performance in more detail, let us consider the cutoff threshold of the ROC curve yielding the smallest distance from the optimal classifier, *i.e.*, the closest point in ℓ_2-norm to the top-left corner in Figure 6, for both the baseline and our proposed method. Using such values we obtain the confusion matrix in Table VI, where we show the classification performance with respect to each of the lower-level vehicle types. As it can be seen, both classifiers performs very well at the extrema of the spectrum, *i.e.* on small vehicles (*CAR* and *SUV*) and on big vehicles (*TRUCK* and *BIGTRUCK*), while performance is worse when trying to classify *PICKUP* from *SMALLTRUCK*. This is likely due to the intrinsic ambiguity in the definition of these classes, as reported in Section III-A. Interestingly enough, the baseline method has better results than our method on 2 out of 7 vehicle types, *CAR* and *SMALLTRUCK*, which more closely match the two types of vehicles used in [3].

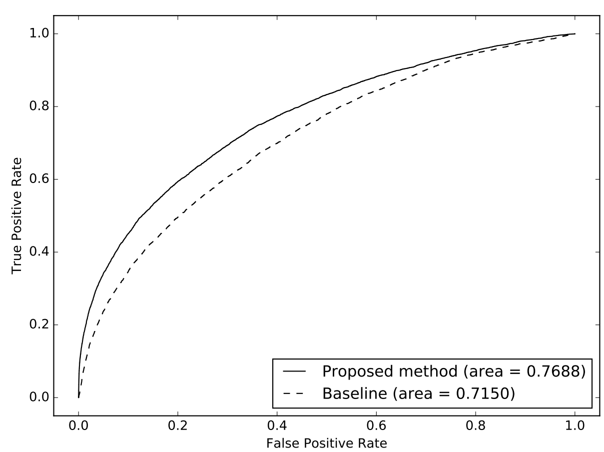


Fig. 5. Classification results when analysing one track at a time

IV. CONCLUSIONS

This paper investigates the yet unexplored problem of vehicle classification from low frequency GPS data. A baseline is created by adapting and fine tuning the approach used by [3], originally proposed for a scenario with higher frequency data. The proposed approach uses features based on travelled distance, speed and acceleration, aggregated over each GPS

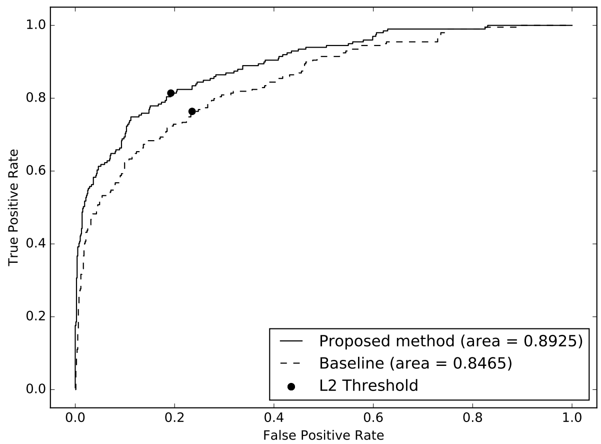


Fig. 6. Classification results when using all the tracks of a vehicle

TABLE VI
CONFUSION MATRIX ON THE MULTIPLE TRACKS PER VEHICLES RESULTS

	Baseline		Proposed method	
	LIGHT-DUTY	*HEAVY-DUTY*	*LIGHT-DUTY*	*HEAVY-DUTY*
CAR	**0.923**	0.077	0.808	0.192
SUV	0.783	0.217	**0.870**	0.130
VAN	0.788	0.212	**0.859**	0.141
PICKUP	0.723	0.277	**0.741**	0.259
SMALLTRUCK	0.435	**0.565**	0.449	0.551
TRUCK	0.205	0.795	0.096	**0.904**
BIGTRUCK	0.053	0.947	0.000	**1.000**

track with several functions. It also exploits the type of travelled roads using a reverse geocoding service. The best set of features is selected automatically via a Recursive Feature Elimination algorithm in a cross-validation framework, where classification is performed via SVM with an RBF kernel.

The experimental results show that the feature selection procedure extracts a set of highly predictive features, letting our method outperform the baseline in terms of area under the ROC curve. The optimal set of features is quite large (69 features) and heterogeneous, but several high level consideration can be drawn. First, total driven distance and road types are within the top 10 features, thus confirming our intuition on their potential value for vehicle classification. Second, some features related to speed are also ranked very high: these can be expected to be more important in a low frequency scenario than acceleration-based ones, but we also speculate that their importance is amplified by exploiting them in conjunction with road types, as our experimental results suggest. Third, some acceleration-related features, albeit less important in the low-frequency scenario, still contribute to classification performance and are worth being included in the set of discriminative features.

Furthermore, we show how the classification of several tracks from the same vehicle can be effectively aggregated, by performing a weighted average over the output of the SVM classifier. Experimental results show that this procedure is able to boost classification performance, both for our method and the baseline.

Several future directions can be envisioned for this work. The dataset could be used to tackle multi-class classification, *e.g.* by considering first a *MID-DUTY* class, and then further up to, potentially, all 7 available vehicle classes. Several additional features could be assessed, such as track tortuosity or GPS altitude. Furthermore, sequence-based kernels could be explored, to investigate if the sequences of GPS points could be directly used without the need of aggregation functions.

REFERENCES

[1] X. Ma and W. E. L. Grimson, “Edge-based rich representation for vehicle classification,” in *Tenth IEEE International Conference on Computer Vision (ICCV’05) Volume 1*, vol. 2. IEEE, 2005, pp. 1185–1192.

[2] S. Gupte, O. Masoud, R. F. Martin, and N. P. Papanikolopoulos, “Detection and classification of vehicles,” *IEEE Transactions on intelligent transportation systems*, vol. 3, no. 1, pp. 37–47, 2002.

[3] Z. Sun and X. Ban, “Vehicle classification using GPS data,” *Transportation Research Part C: Emerging Technologies*, vol. 37, pp. 102–117, 12 2013.

[4] J. H. Wyman, G. A. Braley, and R. I. Stevens, *Field evaluation of FHWA vehicle classification categories*. Maine Department of Transportation, Bureau of Highways, Materials and Research Division, 1985.

[5] M. E. Hallenbeck, O. I. Selezneva, and R. Quinley, “Verification, refinement, and applicability of long-term pavement performance vehicle classification rules,” Tech. Rep., 2014.

[6] R. Benekohal and M. Girianna, “Technologies for truck classification and methodologies for estimating truck vehicle miles traveled,” *Transportation Research Record: Journal of the Transportation Research Board*, no. 1855, pp. 1–13, 2003.

[7] M. Kafai and B. Bhanu, “Dynamic Bayesian networks for vehicle classification in video,” *IEEE Transactions on Industrial Informatics*, vol. 8, no. 1, pp. 100–109, 2012.

[8] G. Leduc, “Road traffic data: Collection methods and applications,” *Working Papers on Energy, Transport and Climate Change*, no. 1, 2008.

[9] G. Xiao, Z. Juan, and C. Zhang, “Travel mode detection based on GPS track data and Bayesian networks,” *Computers, Environment and Urban Systems*, vol. 54, pp. 14–22, 2015.

[10] P. Gonzalez, J. Weinstein, S. Barbeau, M. Labrador, P. Winters, N. L. Georggi, and R. Perez, “Automating mode detection using neural networks and assisted GPS data collected using GPS-enabled mobile phones,” in *15th World congress on intelligent transportation systems*, 2008.

[11] Y. Zheng, L. Liu, L. Wang, and X. Xie, “Learning transportation mode from raw GPS data for geographic applications on the web,” in *Proceedings of the 17th international conference on World Wide Web*. ACM, 2008, pp. 247–256.

[12] A. Bolbol, T. Cheng, I. Tsapakis, and J. Haworth, “Inferring hybrid transportation modes from sparse GPS data using a moving window SVM classification,” *Computers, Environment and Urban Systems*, vol. 36, no. 6, pp. 526 – 537, 2012, special Issue: Advances in Geocomputation.

[13] Y. Wang, Y. Zheng, and Y. Xue, “Travel time estimation of a path using sparse trajectories,” in *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2014, pp. 25–34.

[14] E. Jenelius and H. N. Koutsopoulos, “Travel time estimation for urban road networks using low frequency probe vehicle data,” *Transportation Research Part B: Methodological*, vol. 53, pp. 64–81, 2013.

[15] C. Cortes and V. Vapnik, “Support-vector networks,” *Machine learning*, vol. 20, no. 3, pp. 273–297, 1995.

[16] I. Guyon, J. Weston, S. Barnhill, and V. Vapnik, “Gene selection for cancer classification using support vector machines,” *Machine learning*, vol. 46, no. 1-3, pp. 389–422, 2002.

[17] F. Wilcoxon, “Individual comparisons by ranking methods,” *Biometrics bulletin*, vol. 1, no. 6, pp. 80–83, 1945.