Transportation Research Part C 91 (2018) 176–191

Contents lists available at ScienceDirect

Transportation Research Part C

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

Vehicle classification from low-frequency GPS data with recurrent neural networks

Matteo Simoncini, Leonardo Taccari*, Francesco Sambo, Luca Bravi, Samuele Salti, Alessandro Lori

Verizon Connect Research, Via Giovanni Paisiello, 20, Florence, Italy

ARTICLE INFO

Keywords:
Vehicle classification
GPS
Sequence classification
Recurrent neural networks

ABSTRACT

The categorization of the type of vehicles on a road network is typically achieved using external sensors, like weight sensors, or from images captured by surveillance cameras. In this paper, we leverage the nowadays widespread adoption of Global Positioning System (GPS) trackers and investigate the use of sequences of GPS points to recognize the type of vehicle producing them (namely, small-duty, medium-duty and heavy-duty vehicles). The few works which already exploited GPS data for vehicle classification rely on hand-crafted features and traditional machine learning algorithms like Support Vector Machines. In this work, we study how performance can be improved by deploying deep learning methods, which are recently achieving state of the art results in the classification of signals from various domains. In particular, we propose an approach based on Long Short-Term Memory (LSTM) recurrent neural networks that are able to learn effective hierarchical and stateful representations for temporal sequences. We provide several insights on what the network learns when trained with GPS data and contextual information, and report experiments on a very large dataset of GPS tracks, where we show how the proposed model significantly improves upon state-of-the-art results.

1. Introduction

Inferring the type of vehicles in a road network, a problem typically referred to as *vehicle classification* in the literature, is a fundamental task in several applications, such as surveillance systems, traffic management, emission control, and urban planning (Gupte et al., 2002). Depending on the application, several definitions of vehicle categories can be considered, ranging from very broad ones (*e.g.*, motorcycles, cars, buses), up to fine-grained categorization (*e.g.*, make and model). A standard reference for the definition of classes of vehicles is the 13-category vehicle taxonomy proposed by the Federal HighWay Administration (FHWA) of the United States, which is based on the vehicle weight, length, axles number and axles distances (Wyman et al., 1985). Although the rules have been revised over the years by companies and agencies (Hallenbeck et al., 2014), the FHWA 13 vehicle categories are still used as a standard in many applications.

Traditional approaches for vehicle classification use different types of hardware sensors and classification methods, depending on the application context and the granularity of the desired classification. When physical components can be installed along a road, techniques using fixed-location sensors, such as pneumatic tubes, inductive loop detectors, piezoelectric sensors and Weigh-in-motion

* Corresponding author.
E-mail addresses: matteo.simoncini@verizonconnect.com (M. Simoncini), leonardo.taccari@verizonconnect.com (L. Taccari), francesco.sambo@verizonconnect.com (F. Sambo), luca.bravi@verizonconnect.com (L. Bravi), samuele.salti@verizonconnect.com (S. Salti), alessandro.lori@verizonconnect.com (A. Lori).

https://doi.org/10.1016/j.trc.2018.03.024
Received 12 December 2017; Received in revised form 16 March 2018; Accepted 26 March 2018
0968-090X/

(WIM) systems can be adopted (Benekohal and Girianna, 2003; Harlow and Peng, 2001; Sun et al., 2013). Some of these approaches can even provide the full 13-class classification with high accuracy, in exchange for a high installation cost. For this reason, some less intrusive and expensive classification devices, such as infrared sensors, acoustic sensors and radar sensors, have been introduced at the cost of a lower accuracy and/or coarser classification. Nevertheless, the accuracy of these devices may be affected by traffic and environmental conditions, or human error during installation (Benekohal and Girianna, 2003; Sun et al., 2013).

A different set of techniques involve performing classification from still images or videos, for instance obtained from surveillance cameras (Kafai and Bhanu, 2012; Ma and Grimson, 2005). The number and kinds of vehicle categories varies, depending also on the resolution of the cameras: for instance, vans, taxis, and passenger cars are considered in Ma and Grimson (2005), while sedans, pickups, and vans in Kafai and Bhanu (2012). For image recognition tasks, a huge performance improvement in recent years has been obtained thanks to deep learning approaches based on Convolutional Neural Networks (CNN), such as those presented in Zhou et al. (2016), where the authors classify vehicles into passengers and other vehicles from noisy and/or dark images, and Gebru et al. (2017), where the authors focus on vehicles found in Google Street View images across the United States. In the latter work, a CNN-based classification algorithm is trained on a very large and diverse dataset of almost 350,000 vehicle images and tested on 27,865 examples, obtaining an average class accuracy of 0.67 for the challenging problem of discriminating 11 vehicle types.

For a thorough discussion on advantages and drawbacks of the existing approaches for vehicle classification involving different hardware and techniques, as well as their expected performance, we refer the interested reader to Benekohal and Girianna (2003) and Sun et al. (2013). To give an idea of the performance that can be obtained with different sensors, the authors of Sun et al. (2013) claim that state-of-the-art approaches typically allow an accuracy between 0.80 and 0.95 for 3–5 vehicle classes. We must note, though, that most of the results in the literature are obtained from experiments on small datasets (sometimes as small as a few dozen examples).

Nowadays, the rise of the Internet of Things and connected cars is enabling new ways to sense a vehicle, *e.g.* through the Global Positioning System (GPS) signals it transmits. GPS data are typically produced by either general-purpose mobile devices (*e.g.*, smartphones) or dedicated GPS tracker devices, traditionally installed on commercial or public transport vehicles (*e.g.*, delivery fleets, taxis, ambulances, buses) (Leduc, 2008), but recently also on personal vehicles.[1] When generated by mobile devices, GPS signals are usually used for navigation or geo-localization purposes, hence they exhibit high sampling rates (of the order of one GPS sample per second). The frequency and quality of the data from sensors in modern smartphones allow for their use in several emerging applications, such as driving behavior analysis (Eren et al., 2012; Paefgen et al., 2012; Vlahogianni and Barmpounakis, 2017). When GPS trackers are used, instead, GPS signals are typically used for remote fleet management, vehicle tracking or anti-theft systems, and lower frequency sampling (of the order of one sample per minute) is often 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, as well as for insurance companies (Leduc, 2008). Clearly, the technical and economical advantages come at the cost of accuracy: lower frequency sampling means that information on instantaneous speeds are scarce and that it is harder to infer the true path of a vehicle between two reported positions.

In the context of fleet management software, a strong motivation for tackling the problem of vehicle classification lies in extracting information on the connected vehicles that could not be obtained directly from the GPS devices, and that can be used to improve the user experience of the fleet managers. However, GPS data are still relatively unexplored as clues to tackle this problem. One of the few works that explore vehicle classification from GPS data is Sun et al. (2013), that considers a two-class classification problem, distinguishing between passenger cars and delivery trucks, on a small dataset comprising 52 GPS tracks of passenger cars and 84 GPS tracks of trucks. GPS data used in the paper have a sample rate of 3 s, which is relatively high. The proposed approach entails the use of a Support Vector Machine (SVM) classifier, and the authors conclude that acceleration- and deceleration-based features have a greater predictive power than speed-based features. Closely related to vehicle classification is the problem of transportation mode detection (Bolbol et al., 2012; Gonzalez et al., 2008; Xiao et al., 2015; Zheng et al., 2008), though it is worth remarking that distinguishing between travel modes such as *walk*, *bus*, *train* and *car* is in general easier than the finer-grain detection of vehicle classes, as the former can be solved by using highly discriminative features (such as speed, number of heading changes, number of stops, or distance traveled) that may not be equally effective for vehicle classifications. Very few works in this area, an exception being (Bolbol et al., 2012), consider low-frequency data.

In Simoncini et al. (2016), we addressed for the first time the problem of vehicle classification from low-frequency GPS data, provided by devices installed in commercial fleets for vehicle-tracking purposes. A binary SVM classifier was shown to achieve state-of-the-art performance in distinguishing between light-duty vehicles (*i.e.*, cars, SUVs, vans and light duty pickups, that correspond to classes 2–3 of the FHWA categorization) and larger size vehicles (*i.e.*, heavy duty pickups, small trucks, trucks and big trucks, classes 5–12 of the FHWA categorization). Our method started from a set of low-level features for each point of a GPS track, and then aggregated them at the track level, relying on a wide range of aggregation functions such as mean, median, standard deviation and the like, to obtain over 130 high-level features. The most predictive combination of features was then identified based on a recursive feature elimination procedure (Guyon and Elisseeff, 2003).

In recent years, machine learning approaches based on deep neural networks, commonly known as *deep learning* (Goodfellow et al., 2016), have gained a renewed attention thanks to their success in tasks that involve complex input data like images (He et al., 2015) or sequential or temporal data, such as speech recognition (Deng et al., 2013) and automated machine translation (Cho et al., 2014; Wu et al., 2016), becoming the *de facto* standard technique. What typically makes deep learning effective is the presence of

[1] See for instance the recent Verizon Hum https://www.hum.com.