Autonomous Vehicle Identification based on Car-Following Data

Qianwen Li, Xiaopeng Li*, Handong Yao

Department of Civil and Environmental Engineering, University of South Florida

Keywords: Autonomous vehicle identification; machine learning models; car-following data.

Introduction

Autonomous vehicle (AV) technology holds great potential in enhancing traffic safety, elevating roadway capacity, and assisting AV management and development (Litman, 2020; Woldeamanuel and Nguyen, 2018). Various strategies have been proposed to facilitate AV technology deployment (Arnold et al., 2019; Levin and Boyles, 2015).

While fruitful investigations have been conducted for AVs' benefits, rare efforts have been made to prepare the traffic system for the unprecedented changes induced by AVs. To achieve this goal, identifying AVs from traffic stream is the premise. AV identification technology could extend the video-based traffic surveillance system by adding a dynamic-based AV identification module on the traditional appearance-based vehicle classification framework (Kafai and Bhanu, 2011). With AVs identified, while following AVs, human-driven vehicles (HVs) can drive with caution to reduce crash risk and AVs could reduce the car-following distance for space efficiency. Transportation planners can make appropriate decisions to better handle mixed traffic stream, such as when to open AV dedicated lanes. Further, naturalistic AV field trajectory data is available for assessment. AV safety, mobility, and energy performance can be investigated to recognize AV technology impacts based on realistic AV behavior as compared to simulation data. The yielded insights could be referred to during AV development, i.e., ACC configurations, to promote AV deployment.

Despite such potential, as far as we know, there are no published studies regarding AV identification except a patent assigned to Ford Global Technologies LLC in 2017 (Pilutti et al., 2017). The identification was briefed as collecting related data and then determining. No technical details were introduced.

Motivated by this research gap, this study builds machine learning models to conduct AV identification utilizing vehicle car-following data. Several models are compared including artificial neural network (ANN), long short-term memory network (LSTM), support vector machine (SVM), k-nearest neighbors (KNN), and random forest (RF). Different car-following periods varying from 0.2s to 5s are tested. Identification accuracy of all models stays relatively stable across different periods. The highest identification accuracy is achieved by LSTM with an average of 96.2%, followed by ANN, KNN, RF, and SVM.

Methodology

This section first presents 4 datasets and data preparation, then introduces adopted models.

Datasets

4 datasets are used in this study. The first dataset is HISTORIC data collected by (Yao et al., 2020) with pure HV. The second dataset is Dan's ACC data collected by (Gunter et al., 2020) with pure AV. The third dataset is mixed traffic data collected by this study. The last dataset is part of Open ACC data collected by (Ciuffo et al., 2020) including pure AV and mixed traffic.

Data preprocessing

The above datasets are processed before model training.

Vehicle longitude and latitude are used to compute vehicle location.

1. Linear interpolation is conducted on location to fill up the missing data.
2. Car-following distance is derived based on vehicle location.

3. Moving average smoothing is applied to the collected speed in HISTORIC dataset given the less satisfying GPS accuracy, i.e., 0.28 m/s. Data is resampled to 0.1s to be consistent with other datasets.
4. Vehicle acceleration is calculated as the first-order difference of vehicle speed.
5. Unstable car-following periods are excluded, for example, data at the beginning or end of the test runs.
6. 4 datasets are merged. The car-following distance, preceding vehicle speed, following vehicle speed, and following vehicle acceleration are standardized between -1 to 1 as model inputs.
7. 1 denoting AVs and 0 denoting HVs are set as model output.

After the above data preprocessing, the merged dataset has 1,837,002 data points with an interval of 0.1s.

Machine learning models

This subsection introduces different models. Before model training, data is segmented by the observation period Δt. These observations are randomly shuffled and then separated into 2 subsets with a ratio of 9:1. The greater subset is used for model training and validation. 10-fold cross-validation is conducted. The smaller subset is used for model testing.

Artificial neural network

ANN is the foundation of artificial intelligence. The number of hidden layers N, neurons $\mathrm{N_{neuron}}$, batches $\mathrm{N_{batch}}$, and epochs $\mathrm{N_{epoch}}$ are adjusted to avoid underfitting and overfitting for the best results.

Long short-term memory network

LSTM is a type of recurrent neural network. It is capable of learning order dependence in time series data (Zhang et al., 2019). The number of hidden layers N, dropout layers D, neurons $\mathrm{N_{neuron}}$, batches $\mathrm{N_{batch}}$, and epochs $\mathrm{N_{epoch}}$ and the dropout rate α_k are adjusted to avoid underfitting and overfitting for the best results.

Support vector machine

The SVM model finds the best decision boundary, i.e., decision hyperplane, to separate different classes. The distance from the best hyperplane to the nearest data point of each class is the greatest. Different kernels are tested to produce the best results.

k-nearest neighbors

For the KNN model, we first calculate the Euclidean distance from the query observation to the classified observations. The classified observations are ordered by increasing distance. The class of the query observation is the majority voting of the top k observation classes. k is adjusted for the best model results.

Random forest

The RF model has T decision trees. The maximum depth of each decision tree is $\mathrm{N_T}$. Each decision tree produces each result. The final result is derived as the majority voting of all trees' results. T and $\mathrm{T_N}$ are tuned for the best model performance.

Results

After model tuning, model performance is compared, shown in Table 1. AV identification accuracy of all models stays relatively stable across different observation periods Δt. LSTM produces the highest identification accuracy with an average of 96.2%, followed by ANN (93.7%), KNN (91.5%), RF (91.4%), and SVM (86.9).

Table 1 Model comparison.

Model	Observation period Δt (s)							Average
	0.2	0.5	1.0	2.0	3.0	4.0	5.0	
ANN	92.2	93.4	93.99	93.98	94.01	93.96	93.8	93.7%

LSTM	**96.0**	**96.2**	**96.2**	**96.1**	**96.2**	**96.4**	**96.2**	**96.2%**
SVM	87.3	87.0	86.7	87.2	86.6	86.9	86.8	86.9%
KNN	92.1	91.6	91.8	91.6	91.2	91.0	91.1	91.5%
RF	91.5	91.4	91.7	91.3	91.3	91.8	90.9	91.4%

Conclusion

This study adopts machine learning models to conduct AV identification including ANN, LSTM, SVM, KNN, and RF. 4 car-following datasets involving various AV makes from different manufacturers are merged to build a comprehensive identification model. Different car-following periods varying from 0.2s to 5s are tested. AV identification accuracy of all models stays relatively stable across different periods. The highest identification accuracy is achieved by LSTM with an average of 96.2%, followed by ANN, KNN, RF, and SVM. AV identification can be utilized to benefit roadway users, traffic management, and AV development.

It is noted that not all AVs with different ACC control logics are enumerated here given the resource limit. This study intends to be a methodology demonstration rather than an engineering implementation. Thus, whenever a new AV car-following dataset is available, it can be included in the existing dataset and the presented methodology can be conducted to update the AV identification model.

Acknowledgment

This research is sponsored by National Science Foundation through Grants CMMI #1558887 and CMMI #1932452.

Reference

Arnold, E., Al-Jarrah, O.Y., Dianati, M., Fallah, S., Oxtoby, D., Mouzakitis, A., 2019. A survey on 3d object detection methods for autonomous driving applications. IEEE Transactions on Intelligent Transportation Systems 20, 3782–3795.

Ciuffo, B., Mattas, K., Anesiadou, A., Makridis, M., 2020. Open ACC Database. European Commission, Joint Research Centre (JRC) [Dataset] PID.

Gunter, G., Gloudemans, D., Stern, R.E., McQuade, S., Bhadani, R., Bunting, M., Delle Monache, M.L., Lysecky, R., Seibold, B., Sprinkle, J., 2020. Are commercially implemented adaptive cruise control systems string stable? IEEE Transactions on Intelligent Transportation Systems.

Kafai, M., Bhanu, B., 2011. Dynamic Bayesian networks for vehicle classification in video. IEEE Transactions on Industrial Informatics 8, 100–109.

Levin, M.W., Boyles, S.D., 2015. Effects of autonomous vehicle ownership on trip, mode, and route choice. Transportation Research Record 2493, 29–38.

Litman, T., 2020. Autonomous vehicle implementation predictions: Implications for transport planning.

Petrović, Đ., Mijailović, R., Pešić, D., 2020. Traffic Accidents with Autonomous Vehicles: Type of Collisions, Manoeuvres and Errors of Conventional Vehicles' Drivers. Transportation research procedia 45, 161–168.

Pilutti, T.E., Rupp, M.Y., Trombley, R.A., Waldis, A., Yopp, W.T., 2017. Autonomous vehicle identification.

Woldeamanuel, M., Nguyen, D., 2018. Perceived benefits and concerns of autonomous vehicles: An exploratory study of millennials' sentiments of an emerging market. Research in Transportation Economics 71, 44–53.

Yao, H., Li, Q., Li, X., 2020. A study of relationships in traffic oscillation features based on field experiments. Transportation Research Part A: Policy and Practice 141, 339–355.

Zhang, X., Sun, Jie, Qi, X., Sun, Jian, 2019. Simultaneous modeling of car-following and lane-changing behaviors using deep learning. Transportation research part C: emerging technologies 104, 287–304.