Driver intention prediction of stop length using LSTM

Joerg Christian Wolf
Innovation and Engineering Center California
Volkswagen Group of America, Inc.
Belmont, CA, United States
ORCID: 0000-0001-5695-3490

Tianxing Yang
Innovation and Engineering Center California
Volkswagen Group of America, Inc.
Belmont, CA, United States
ORCID: 0000-0002-4027-7640

***Abstract*—Understanding the intentions of a driver at an intersection is a useful property for a variety of driver assistance functions ranging from engine management to safety. This paper shows how a time series prediction can be modelled to classify the intended stop length with high accuracy in a real-world driving situation. Stop length is defined by how many seconds a vehicle waits at a position before continuing its journey. An LSTM was activated when the vehicle comes to a stop and processes the last few seconds of data signals from the vehicle. Many possible input signals have been explored. It was shown that specifically the time series of the brake pressure and the steering wheel angle contribute the most to an accurate prediction.**

Keywords—CAN, LSTM, intention prediction

I. Introduction

Intention prediction is an enabler for a variety of driver assistance functions. The knowledge about what the driver and vehicle will do in the next few seconds can improve safety when embedded in an ADAS (Advanced Driver Assistance Systems) [1] or improve the engine power management [2]. Many different models have been proposed to model driver behavior and predict the next action, such as a lane change. For instance Trivedi's group [1][9][13] used a relevant vector machine (RVM) to predict turns and lane changes from head movement, lane position, surrounding vehicle positions and pedal sensors. In corporation with Volkswagen, they were able to access the output of these driver assistance systems to create a comprehensive context for the prediction. Hou et al. [7] and Berndt et al. [8] use a Hidden Markov Model (HMM) to model lane change intention. Hou et al. [7] uses basic signals such as steering wheel angle velocity and lateral acceleration as input signals to the HMM.

While driver intention prediction may include a variety of sensors including driver monitoring or even monitoring of foot gestures [9], it is more cost effective to use signals from the vehicle data bus such as CAN (Controller Area Network). In previous work we found that CAN data contains distinct features of driving characteristics, see Fugiglando et al. [4] . In another collaboration it was found that it is possible to predict CAN bus signals of a vehicle with Recurrent Neural Networks (RNNs), see Hallac et al. [3]. The philosophy here is to take all available CAN signals indiscriminately and let an RNN determine which signals are important through its weights. This can be thousands of signals in modern vehicles. Hallac et al. [3] has shown that through RNNs it is possible to predict maneuvers at their onset. We believe that compared to SVM, RVM or HMMs an RNN is able to model temporal aspects of human activities better through its deep architecture.

In this paper, we are going to create a model that predicts how long a driver intends to stop a vehicle. This is an important property for the vehicle's power management [5][6][16] or one can imagine in-car advertisement [10] or other services that the driver may get depending on the length of the stop. Stop length prediction can also be useful for intelligent public transportation, such as bus arrival times and bus stop information. Liu et al. [11] suggests a system based on an LSTM or Long Short-Term Memory (LSTM) in combination with geolocation data.

While there are some patents [6] roughly covering the topic stopping or maneuver intentions, the authors are not aware of any research specifically addressing the stop length prediction problem and an LSTM solution. In the methodology section, we describe the collected data, the model and input signals. Then the performance of the network is presented in the results section. The paper finishes with a discussion and conclusion section.

II. Methodology

A. *Data Collection*

Sixteen vehicles belonging to Volkswagen staff in California have been equipped with CAN signal collection

This work was funded by the Volkswagen Group.
We thank Mapillary AB for the traffic sign data.

devices with their consent. The vehicles were used for daily commutes and regular driving over a period of approximately 6-8 months. The vehicle journeys accumulated to 14310 kilometers. The drivers were not biased since they did not receive any instructions on how to drive. The driving data contains complete trips that include urban, sub-urban, highway driving and parking mainly around the California Bay Area. After initial filtering for impact we reduced the CAN signal set to 11 signals that include accelerator pedal setting and gradient, road slope, distance to vehicle in front, brake pressure and gradient, steering wheel angle and some signals on the vehicle's dynamics. For experimentation, we later added other input signals such as road signs or traffic conditions. The signals were recorded at different sampling rates, up to 50 Hz, depending on the source. In a post-processing step the signals were sampled to be all at 10 Hz. The sampling rate was chosen to be high enough to capture the driver's motion on the pedals and steering wheel.

B. *Model*

The aim is to create a model that can distinguish between short and long vehicle stops. Depending on the application, other classifiers or a regression model are also possible. We can identify the length of stops from the vehicle's velocity on the data bus and can therefore automatically label the onset and length of a stop accurately. First, the distribution of the length of stops was visualized to see what kind of stops are most frequent.

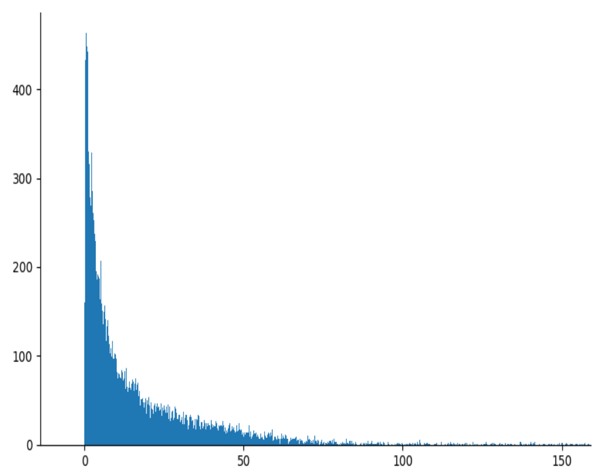


Fig. 1. Histogram of frequency of stops divided into 0.1 second buckets

The distribution in figure 1 shows that the most frequent vehicle stops are less than one second long and most stops took less than 3 seconds. Let us define a "short stop" as being shorter than 2 seconds for modelling. The chosen limit depends on the application of the model. For example for engine management, a 2-second limit is a useful property. Longer stops are less frequent. Short stops are common when maneuvering in a car park or property driveway. In the United States, short stops can also occur at a right turn on red or stop sign. Short stops can also occur when inching forward in a queue of vehicles. The vehicles of this study are mostly automatic transmission vehicles.

For this machine-learning approach, it is important to first have an overview by presenting this histogram and making sure there are enough examples of different length of stops in the data set. The histogram shows that there are thousands of stops with a length of less than 50 seconds, which is a sufficiently sized dataset.

So once a vehicle stops (speed = 0) we can select the last t_w seconds of signals as an input set that lead up to the stop. We experimented with how many seconds of the approach to the stop are relevant for a good prediction, see hyper-parameter table 1.

After we have gathered and preprocessed the data, the next step is to build the neural network that uses the LSTM technique to predict if a long or short stop is coming up. The LSTM (which stands for "Long Short-Term Memory") [20] network is a special kind of RNN (recurrent neural network), which aims to address the traditional RNN's issues with recognizing long-term dependencies within the streams of information being processed. Since not every single piece of prior information has equal importance in predicting upcoming information, the LSTM network is designed to identify what specific features have greater importance in determining the outcome, and remember the features that have long-term importance in the information being processed.

For this project, in real life situations, whether the car will have a long stop or a short stop is dependent on many factors. This is also reflected by the changes in the motion patterns of the car prior to comping to a stop. The car's motions are influenced by a number of factors, including traffic signals/signs, the intended direction of the car after passing through the intersection (straight/left/right/U-turn), the amount of traffic on the road, the density of intersections, etc. Since there are so many factors that may influence the car's motion, some of which take place quite a "long" time (a few seconds) before the stop of the car, a neural network must be able to recognize which factors are playing more important roles than others in determining the upcoming stop. The network must also determine which specific features have relatively greater long-term influences than other features. This is a primary factor that the LSTM neural network architecture is suitable for.

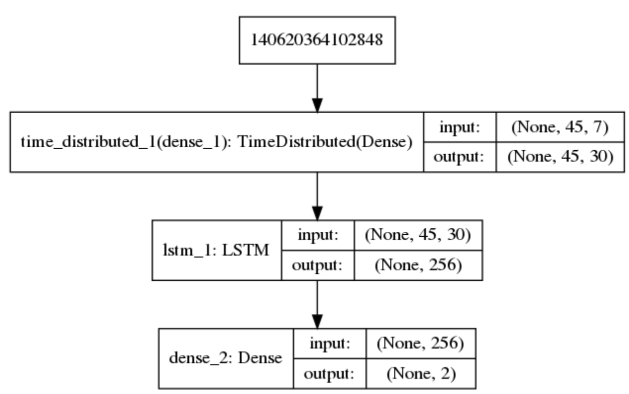


Fig. 2. The LSTM model

We constructed the training model using Keras and Tensorflow [14], see figure 2. We first added a TimeDistributed layer to divide the data by each single stop of the car, keeping a short period of data before the car stops each time (5 seconds). The TimeDistributed layer is a wrapper layer that applies the Dense layer independently to each batch of the input data, creating an intermediate layer of nodes to be fed into the LSTM layer. The LSTM layer plays the core role of determining which input signals have greater importance in determining whether the result would be a short or long stop. Then the final Dense layer outputs a 2-D vector as the predicted result, corresponding to either the short stop or the long stop, which uses the softmax activation function.

For better training outcomes and avoiding overfitting, we have applied the keras default kernel regularization technique on the TimeDistributed layer with $l_1 = 0.1$ (uses absolute) and $l_2 = 0.1$ (uses square). We have also added the recurrent dropout parameter 0.4 and the dropout parameter 0.4 on the LSTM layer.

The output of the model are 2 nodes that act as classifiers for a short stop or long stop.

C. *Input Signals*

The signals we have tried are: distance to car in front; brake pressure gradient; brake pressure; angular velocity; longitudinal acceleration; lateral acceleration; wheel velocity; steering wheel angle; accelerator pedal position gradient; accelerator pedal position; GPS heading; GPS latitude; GPS longitude; slope of the road; traffic control sign ahead.

The car's velocity signal is used to determine if the vehicle has stopped. It provides the ground truth for the predictions and is used for auto-labelling the stops.

D. *STOP / YIELD Sign Identification*

To investigate the correlation between stop lengths and these traffic signs, we have gathered the traffic sign data of the Belmont, CA area from Mapillary AB. Since both the traffic sign data and the car operation data involve the latitude, longitude, and facing angle, we can find out whether the driver is facing the stop sign during the driving, and calculate the distances between the car and the STOP / YIELD sign over the period in which the car driver observes the sign, slows down, and comes to a stop. We can limit the range of observation to around 45 degrees to the left and right in front of the driver, since this is the approximate range that the driver can observe road conditions ahead and react to the conditions.

.

III. RESULTS

We have trained the model with different hyper parameter combinations. The plot below shows the weight decay after running 100 epochs of training.

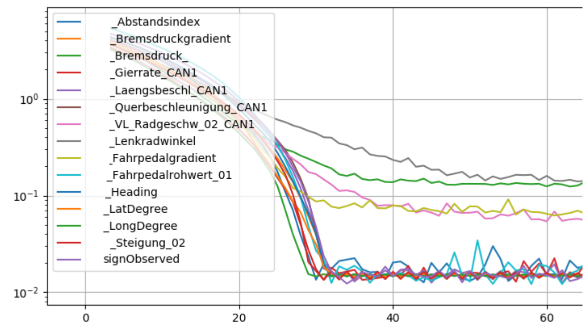


Fig. 3. weight decay

Weights of the first layer determine the importance of the signal going into the neural network. These are usually reducing as the training goes on and can be enhanced with L2 regularization. Some input weights stabilize at a much higher value, which means they are relevant for the network to perform well. Four CAN signals have been found to have significantly greater importance than other signals:

- Steering wheel angle
- Brake pressure
- Brake pressure gradient
- Accelerator pedal position gradient

In another data-set (that was not part of this study), an attempt was made to test the influence of the traffic conditions. The amount of traffic (low to congested) that was present at this section of road in the last 5 minutes was normalized and quantified as 3 inputs to the LSTM: minimum, maximum and mean traffic. The weight decay showed that these additional 3 inputs did not have an significant effect on the accuracy and the input weights had a low value, like at the bottom of figure 3.

During one training epoch, the thousands of stops where divided into batches with 512 samples of stops each. The overall data was randomly split (per vehicle stop) into 75% for training and 25% for validation. The resulting receiver operating characteristic (ROC) curve is shown in figure 4.

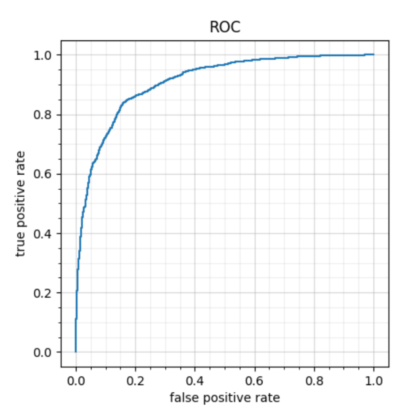


Fig. 4. ROC

It can be seen that after around 60 epochs the weights of different CAN signals have stabilized. We have also tried out the performance of our model using different hyper-parameters, see table 1.

TABLE I. HYPER PARAMETERS

Number of Epochs	*Time Steps*	*Testing Accuracy (%)*
60	250	84.1
60	200	84.2
60	100	82.0
60	50	82.2
100	250	83.0
100	200	83.1
100	100	81.0
100	50	81.2

We were also interested to see if a stop or yield sign is a good indicator for the length of a stop so we counted the number of long/short stops and if there was a sign or not in table 2.

TABLE II. TYPE OF STOP AND IF IT HAS A TRAFFIC CONTROL SIGN

Type of Stop	*No Sign*	*Sign*
Short Stop (< 2 sec)	12%	9%
Long stop	21%	58%

One important fact to be noted is that whenever a car encounters a STOP or YIELD sign, most of the times it turns out to be a long stop. When encountering a sign, the current sampling data shows that among the samples that are associated with STOP / YIELD signs, 9% of the samples are short stops, while 58% are long stops. This significant difference implies a strong correlation between STOP / YIELD signs and long stops.

IV. GENERAL OBSERVATIONS

Since short stops are frequent when maneuvering in a car park, then this can correlate with a low speed and large steering-wheel movements.

It turns out that only a few seconds ($t_{w=}$5 sec) leading up to a vehicle stop are relevant for predicting if it going to be a long or short stop. Longer time windows lead only to very small improvements, which means that they are not so relevant. The time window length roughly corresponds to the time it takes for a driver to slow down when approaching an intersection, Therefore one could conclude that drivers usually don't take any action earlier than 5 seconds before a stop. Min et al. [12] modelled the deceleration in order to predict its shape and found that there is an initial increase of braking, then an adjustment phase followed by a termination phase where the deceleration is reduced again.

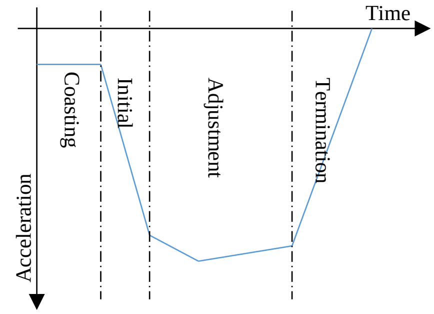


Fig. 5. According to Min et al. 12], the braking can be partitioned into 4 sections: *Coasting*, when the foot moves from accelerator to brake pedal, *Initial* braking section with a jerk, *Adjustment* section when the driver tries to maintain a constant deceleration and finally the *termination*, where the braking ends.

Min [12] found that depending on if the driver is approaching a speed bump, curved road or traffic signal the shape of the deceleration was different. We can agree with this observation. However, we can see from our dataset and model weights that the type of upcoming stop is not a reliable indicator that can contribute to an accurate stop length prediction. In future research, it may be interesting to visualize the patterns and parameters and their correlation with different types of stops.

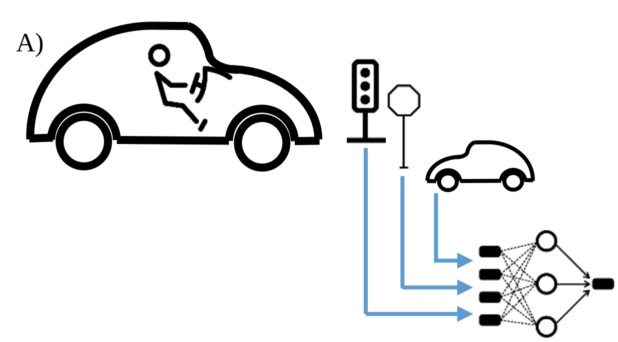
V. PERCEPTION WITH DRIVER IN THE LOOP

While experimenting with different input signals to the LSTM we can broadly categorize into two types of sources:

A) Observation from the environment

B) Action of the driver

Our intuition was to start with signals from the environment, such as road signs, vehicles waiting in front, traffic light information [15], shown in figure 6.



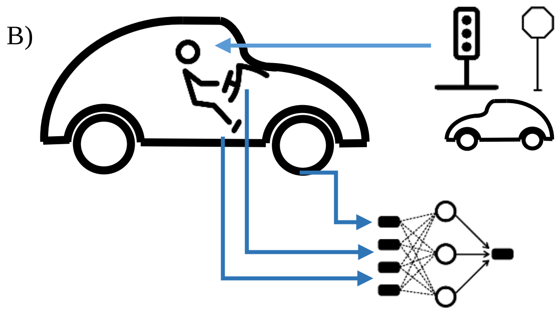


Fig. 6. Perception of input signals of the LSTM through the environment (A) or from the driver (B) who sees the environment

We found that using the signals from the driver, like in figure 6-B) gave the best accuracy and outperforms signals from the environment when looking at accuracy and weight decay. So why is that ?

Information from the environment, such as the position of other vehicles in traffic [16] and road signs all have some relevance to making a decision on how long to stop. However, depending on the situation some are more relevant. Furthermore, the vehicles used for this study have driver assistance features but do not have all information available on the vehicle bus that may be relevant for the decision-making. The question on what is relevant is a cognitive filtering process. It is guided by attention. Cognitive modeling and traffic psychology [17] bring insights to how a driving task can be modelled. Typically a perception and attention system is coupled with a closed loop control system that controls the more automated task of maintaining vehicle speed and trajectory. Moving objects get the drivers attention. Especially when being on a potential collision course the attention span extends. David Lee [18] modeled the braking behavior as being influenced by the time-to-collision.

So in essence, the driver knows what kind of stop is coming up and has an expectation how long it will be. This is represented in for example how much brake pressure is applied when the vehicle is coming to a stop or how the steering is handled. The fact that the LSTM can learn to predict the length of the stop is evidence that drivers treat long stops differently with their actions. Those actions are easier to measure than the outside environment, since for driving, all a driver can do is use the pedals and the steering wheel, thus making the model in figure 6-B) more successful.

The result of 84.2% accuracy was achieved with a randomly mixed set of all drivers based on thousands of stops. We also experimented with using data only of a single driver. While this may be able to improve the accuracy. The data likely to be driver dependent. Depending on the application of the model, it is possible to consider training the network with data of a single driver. Hallac et. al. [19] showed that it is possible to identify a driver based on those CAN signals. This is also an indication that steering and acceleration is driver dependent. The accuracy is high enough to consider applying the model for vehicle's power management like in [5][6][16], since it is not energy efficient to turn off the engine for a very short stop.

VI. Conclusion

This paper has shown that it is possible to predict the length of a stop with an LSTM using CAN signals as input that monitor the driver. The concept of monitoring the driver's actions rather than the environment was introduced. The driver's attention and actions act like a filter that enables a high accuracy.

References

[1] A. Doshi, B. Morris and M. Trivedi, "On-road prediction of driver's intent with multimodal sensory cues" in *IEEE Pervasive Computing*, vol. 10, no. 3, pp. 22-34, July-September 2011.

[2] R. Wang and S. M. Lukic, "Review of driving conditions prediction and driving style recognition based control algorithms for hybrid electric vehicles," *2011 IEEE Vehicle Power and Propulsion Conference*, Chicago, IL, 2011, pp. 1-7.

[3] D. Hallac, S. Bhooshan, M. Chen, K. Abida, R. Sosič and J. Leskovec, "Drive2Vec: Multiscale State-Space Embedding of Vehicular Sensor Data," *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*, Maui, HI, 2018, pp. 3233-3238.

[4] U. Fugiglando, E. Massaro, P. Santi, S. Milardo, K. Abida,R. Stahlmann,F. Netter, C. Ratti, "Driving Behavior Analysis through CAN Bus Data in an Uncontrolled Environment," in *IEEE Transactions on Intelligent Transportation Systems*, vol. 20, no. 2, pp. 737-748, Feb. 2019.

[5] T. Schoen and F. Riegert, "Prognostizieren einer voraussichtlichen Haltezeit für ein Start-Stopp-System eines Kraftfahrzeugs", German Patent DE 10 2017 200 602 A1, Jul. 19, 2018

[6] A. Piggott, "Intelligent engine idle stop logic", US Patent US20120095670A1 Oct. 13, 2010

[7] Hou, Haijing & Jin, Lisheng & Niu, Qingning & Sun, Yuqin & Lu, Meng. "Driver Intention Recognition Method Using Continuous Hidden Markov Model" in *International Journal of Computational Intelligence Systems*. May 2011. 386-393. 10.1080/18756891.2011.9727797.

[8] H. Berndt, J. Emmert and K. Dietmayer, "Continuous Driver Intention Recognition with Hidden Markov Models" in *11th International IEEE Conference on Intelligent Transportation Systems*, Beijing, 2008, pp. 1189-1194, doi: 10.1109/ITSC.2008.4732630.

[9] C. Tran, A. Doshi, M.M. Trivedi, "Modeling and prediction of driver behavior by foot gesture analysis" in *Computer Vision and Image Understanding*, vol. 116, no 3, 2012, pp 435-445, ISSN 1077-3142

[10] H. Aksu, L. Babun, M. Conti, G. Tolomei and A. S. Uluagac, "Advertising in the IoT Era: Vision and Challenges," in IEEE Communications Magazine, vol. 56, no. 11, pp. 138-144, November 2018, doi: 10.1109/MCOM.2017.1700871.

[11] D. Liu, J Sun, S Wang "BusTime: Which is the Right Prediction Model for My Bus Arrival Time?" arXiv:2003.10373 , Mar 2020

[12] K. Min, G. Sim, S. Ahn, M. Sunwoo, K. Jo, "Vehicle Deceleration Prediction Model to Reflect Individual Driver Characteristics by Online Parameter Learning for Autonomous Regenerative Braking of Electric Vehicles" in *Sensors* 2019, 19, 4171.

[13] B. Morris, A. Doshi and M. Trivedi, "Lane change intent prediction for driver assistance: On-road design and evaluation," 2011 IEEE Intelligent Vehicles Symposium (IV), Baden-Baden, 2011, pp. 895-901, doi: 10.1109/IVS.2011.5940538.

[14] M. Abadi, et al, "TensorFlow: Large-scale machine learning on heterogeneous systems", white paper 2015. from https://tensorflow.org.

[15] J. C. Wolf, J. Ma, B. Cisco, J. Neill, B. Moen, C. Jarecki "Deriving Signal Performance Metrics from Large-Scale Connected Vehicle System Deployment" in the "Transportation Research Record","(Journal of the Transportation Research Board)" doi: 10.1177/0361198119838520 , Apr 2019

[16] S. Engel "Verfahren zum Betrieb eines Start-Stopp-Systems eines Kraftfahrzeugs und Kraftfahrzeug" European Patent, EP2592261A1, 15 May 2013, Issue 2013/20

[17] J Pekkanen, O Lappi, P. Rinkkala, S Tuhkanen, R. Frantsi H Summala, "A computational model for driver's cognitive state, visual perception and intermittent attention in a distracted car following task." Royal Society Open Science. 10.1098/rsos.180194, 2018

[18] D. N Lee. "A theory of visual control of braking based on information about time-to-collision", Perception Vol 5,pp. 437–459, 1976 , doi:10.1068/p050437

[19] D. Hallac, A. Sharang, R Stahlmann, A. Lamprecht, M. Huber, M. Roehder, R. Socic, J. Leskovec, "Driver Identification Using Automobile Sensor Data from a SingleTurn", 2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC), pp. 953–958, 2016.

[20] Sepp Hochreiter, Jürgen Schmidhuber; Long Short-Term Memory. Neural Comput 1997; 9 (8): 1735–1780. doi: https://doi.org/10.1162/neco.1997.9.8.1735