2018 21st International Conference on Intelligent Transportation Systems (ITSC)
Maui, Hawaii, USA, November 4-7, 2018

978-1-7281-0323-5/18/$31.00 ©2018 IEEE

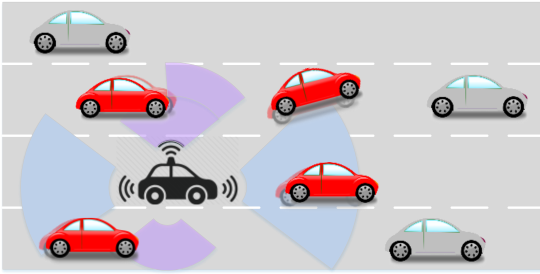

Fig. 1. Limited observation of SVs with occlusion and noise

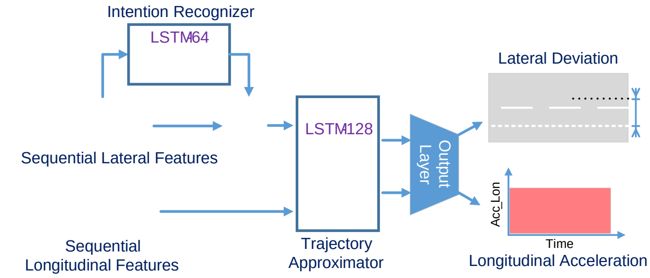

Fig. 2. Proposed system architecture

methods in fields from image classification [15] to natural language processing [16]. Due to the unique neural network structure, recurrent neural network (RNN) is widely used to learn the temporal structure of time-series data.

In particular, the long short term memory (LSTM) architecture has been successfully applied to analyze the temporal structure underlying in text, speech, and financial data [17], as well as lateral position prediction [18] and driver intent prediction [19] in the field of intelligent driving. Reference [20] directly applied LSTM to predict future trajectory over occupancy grid. Although the occupancy grid helps to deal with the prediction uncertainty, the method sacrifices prediction accuracy in terms of position as the grid is 10m long. Reference [21] proposed an LSTM network for highway trajectory prediction that leveraged the information of surrounding vehicles around the target vehicle. The main downside, however, is the infeasibility of observing these information for a target vehicle which is already a surrounding vehicle of the ego vehicle due to sensing occlusion. Again, similar to prototype trajectories, these studies are not capable of dealing with different road layouts in a universal way as it is difficult to tell a lane change behavior from driving along a curved road, thus misleading the prediction results.

In contrast to the existing methods, this paper proposes a more pragmatic and geography-adaptive method which is first to estimate the maneuver intention of the driver and then to predict the successive kinematic states in future horizon to correspond with the possible execution of the identified maneuver. In our work, we employ two LSTMs to recognize high-level driver intention as well as to understand low-level complex dynamics of vehicle motion. The proposed trajectory prediction system inputs the sequentially transformed coordinates of the surrounding vehicles obtained from the sensor measurements to the LSTMs and produces the YHKLFOH¶V future locations in the next 5 seconds. The method is developed based on naturalistic driving data and shows better prediction accuracy over some existing methods.

The rest of this paper is organized as follows. In Section II, the problem is formally defined. In section III, the proposed method and its related techniques are introduced. The dataset used and the training method employed are described in Section IV. In Section V, the results are provided and the paper is concluded in Section VI.

II. PROBLEM STATEMENT

To achieve better accuracy in decision making and path planning modules of an autonomous vehicle, it is beneficial to address the problem of predicting future trajectories of surrounding vehicles on a highway using historic but limited observation data from on-board sensors. Different from the work in [20] where vehicle trajectories are represented in the form of occupancy grid and turned the prediction task into a classification problem, we consider it as a regression one whose outputs are supposed to be as accurate as actual values.

More formally, our goal is to train a predictor for future trajectory of sequential outputs $Y = \{y_t\}_{t\in T_{post}, y\in \mathcal{O}}$ based on a set of observable feature inputs $X = \{x_t^k\}_{t\in T_{prev}, x^k\in \mathcal{F}}$, where $T_{prev} = \{-t_{pres}, \ldots, 0\}$ and $T_{post} = \{1, \ldots, t_{post},\}$ are respectively the time intervals of the historical input and future predictions, $\mathcal{O}$ is the target outputs and $\mathcal{F}$ is the feature set acquired and calculated simultaneously. For $x^k \in \mathcal{F}$ and $t \in T_{post}$, we denote x_t^k by the value of feature x^k observed t time steps earlier. Similarly, we denote y_t by the value of output t time steps in the future.

In this paper, future trajectories of surrounding vehicles to be predicted only depend on information observed from the ego vehicle which is usually partially observable due to sensor limitations and object occlusions. In contrast to [21] which used information of vehicles around a target surrounding vehicle to the ego vehicle, our assumption on the data availability is more pragmatic for real driving scenarios. In order to overcome the challenges of information limitation, data imbalance between lane keeping and other driving maneuvers, and model generalization, a novel framework based on intention-aware LSTM network is proposed and presented in Section III.

III. PROPOSED METHOD

A. System Architecture

In this paper, the proposed deep RNN network is presented in Fig. 2. It mainly consists of two LSTM networks, one with a layer of 64 cells and the other with a layer of 128 cells, followed by a dense (fully connected) output layer containing as many neurons as the number of outputs. In this architecture, the role of the first LSTM layer is to recognizH GULYHUV¶ intentions, e.g. lane keeping and lane change, and transform the lateral features while the second LSTM layer is to extract meaningful representation of the sequential inputs and thus produce future trajectories with help of the dense layer and the embedded semantic understanding of driving intentions.

Compared to the existing studies [20][21], this architecture enjoys several advantages. One is that the semantic understanding of driving intention can be used as instructive information for inferring VXUURXQGLQJ YHK motion. Another is that upper and lower boundaries for lateral trajectory predictions can be generated based on prior

$$\hat{v}_y(t) = \hat{v}_y(t-1) + \hat{a}_y(t) * \delta t \qquad)$$

$$\hat{y}(t) = \hat{y}(t-1) + \hat{v}_y(t) * \delta t \qquad)$$

where δt is the time step.

Since the lateral position is bounded, the lateral deviation over the prediction horizon $\{\hat{x}_{dev}(t)\}_{t\in T_{post}}$ are directly used for the output. Then the lateral position can be calculate as

$$\hat{x}(t) = \hat{x}_{dev}(t) + \hat{x}_{targ} \qquad)$$

where $\hat{x}_{targ}$ is the local lateral position of the centerline of the inferred target lane based on the intention recognition output from the first LSTM layer.

IV. Data and Trainning

A. Data

The dataset used in this paper is from the Next Generation Simulation (NGSIM) [22]. Collected and published by the US Federal Highway Administration in 2005, the NGSIM is one of the largest open datasets of naturalistic driving and has been widely studied in the literature, e.g., [21][23][24].

More specifically, the area of interest is the I-80 freeway in Emeryville, California, of which the covered segment is approximately 500m in length and 6 lanes (3.66m or 12ft each) in width (see Fig. 2). The 45-minute trajectory data were collected from 4:00pm to 4:15pm and from 5:00pm to 5:30pm, reflecting different traffic characteristics during transitional and congested traffic period, respectively.

The dataset contains more than 5000 trajectories of individual vehicles, with a sampling rate at 10 Hz. Each sample in one trajectory includes the information such as instantaneous speed, acceleration, longitudinal and lateral positions, vehicle length, and vehicle type. The local coordinates is set at the down-left point of the study area, where x is the lateral position of the vehicle relative to the leftmost edge of the road, and y its longitudinal position to the entry edge. Furthermore, a total of 914 successful lane changes were identified automatically per SAE J2944 [25], while the lane change to the left totaled 694.

B. Training

The proposed network is trained using a window of 50 time steps, representing 5s past observations. Such window is updated every second at 10 Hz. To increase training efficiency and avoid back propagation related issues, input data are grouped by batches of 100 and shuffled within batches.

For the first LSTM layer, it can be interpreted as a multi-class classification problem. Driving intentions, namely lane keeping, left lane change and right lane change, are encoded into a 1x3 vector with one hot technique when training. A softmax layer is added to the LSTM output with the linear transformation of hidden state h_t followed by the softmax function

$$P(y=j|h_t) = \frac{e^{{h_t}^T w_j}}{\sum_{i=1}^{k} e^{{h_t}^T w_i}} \quad for\ j = 1, \dots, k \qquad)$$

where y is the driving intention prediction; w_i are the weights related to each intention category; $k = 3$ is the number of intention categories.

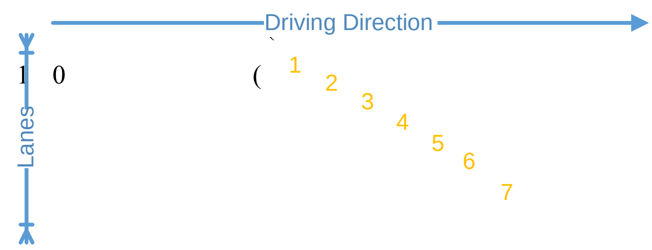

Fig. 4. Birdview of naturalistic traffic recorded on I-80 freeway

For the second LSTM layer, defined features are directly taken as inputs and the loss function is set as the quadratic form of difference between the network predictions and the actual values. L_2 loss is calculated and weights are updated through back propagation through time (BPTT) .

V. Results

This paper explored the long-time horizon (up to 5s) prediction of future vehicle trajectories for different driving intentions based on LSTMs. Samples are from the NGSIM I-80 dataset (70% for training, 30% for validation). The training is performed on GPU using the TensorFlow for 5 epochs. The initial learning rate is set to 1.0 and gradually decreases until the validation error stops to improve.

For the test, the input features of each trajectory are directly fed to the network, without additional process, e.g., grouping the data by time windows and priori selection of specific trajectory segments. The proposed network takes features extracted from past trajectory as input, recognize the driver intention on a probability basis and simultaneously predict future trajectory as output. Motion predictions of three different vehicles (identified as 831, 967 and 928) are presented in Fig. 5, with time difference of 2s. Their trajectories are represent by solid lines, dashed lines and dotted lines, respectively, with predicted trajectories in red and real ones in yellow. The proposed method analyzes and adapt itself to individual driving characteristics as the predicted trajectories differ from each other. Compared through Fig. 5 (a) to (c), a lane change action of vehicle 831 is recognized in time and in advance on a probability basis in the background system. With the help of intention recognition, the network is able to make human-like judgment of the motion pattern of surrounding vehicles, thus providing a good guidance for the low-level trajectory predictor.

For each individual trajectory, the Root Mean Squared Error (RMSE) between the predicted and the actual value is then computed to assess the learning performance of the proposed network as well as the generalization capability over different drivers. Results using other LSTM-based methods are introduced for comparison purposes. One is proposed by [20] which predicts future trajectory over occupancy grid. The other one is proposed by [21] which leverages information of vehicles around target vehicles.

Our proposed model provides the best overall results for longitudinal position prediction and similar results to the best prediction for lateral position. Note that [21] takes much more vehicle information. Therefore, compared to other models, the proposed model extracts semantic information of driving intention which helps considerably to narrow down prediction boundaries even though the observation is limited. As for the