Vehicle trajectory prediction based on LSTM network

Zhifang Yang
school of electrical and information engineering
Wuhan Institute of Technology
Wuhan, China
wit_chuangxin@163.com

Dun Liu
school of electrical and information engineering
Wuhan Institute of Technology
Wuhan, China
1209986493@qq.com

Li Ma
school of electrical and information engineering
Wuhan Institute of Technology
Wuhan, China
983677559@qq.com

2022 International Conference on Artificial Intelligence and Computer Information Technology (AICIT) | 978-1-6654-5087-4/22/$31.00 ©2022 IEEE | DOI: 10.1109/AICIT55386.2022.9930177

***Abstract*—In a complex traffic environment, predicting the trajectory of surrounding vehicles in the driver's line of sight can greatly reduce the possibility of various traffic accidents and play an auxiliary role in the driver's decision making. The motion of predicted vehicles is constrained by the traffic environment, that is, the motion of adjacent vehicles and the relative spatial positions between vehicles. This paper mainly studies the behavior prediction of vehicles on the expressway. Based on the social convolutional pooling LSTM network (CS-LSTM), a CS-LSTM network with an attention mechanism is proposed, which assigns different weights to the fusion features and improves the accuracy of the trajectory prediction of surrounding vehicles. This article evaluates the model on a publicly available NGSIM dataset. The results show that the proposed algorithm is more accurate than other algorithms in predicting the future trajectory of vehicles.**

Keywords—Long Short-Term Memory, trajectory prediction, attention mechanism

I. INTRODUCTION

In order to ensure the safe and efficient driving of drivers in complex traffic, predicting the trajectory of surrounding vehicles can help guide drivers to change lanes, overtake, decelerate, and allow other vehicles to merge in advance [1], which requires the system to be able to It has the ability to reason about the future motion trajectories of surrounding vehicles. This can be seen in existing path planning algorithms [2], which require a reliable estimate of the vehicle's future trajectory.

There are a large number of latent variables in vehicle trajectory prediction due to the driver's ultimate purpose and the differences in the driver's driving style. Due to the driver's driving situation, the vehicle trajectory is in a non-linear state over a long period of time; in addition, the driver's behavior is multimodal, that is, in the same scene, the driver can make one of several decisions; The interaction between the trajectories will change the motion trajectories [3], so predicting the trajectories of surrounding vehicles is a challenging problem.

Vehicle trajectory prediction is divided into physical model-based, behavior-based and interactive perception-based models [4]. The method based on the physical model only considers the speed of the object, the state of the acceleration object, etc. to predict the trajectory, such as Bahram M, etc., using the physical model combined with the intention prediction classification, using the heuristic learning method, which simplifies the operation of the model [5]. Batz et al. expressed the state parameters of the vehicle through the Gaussian distribution method, and finally converted it into the road environment location information to obtain the predicted vehicle trajectory behavior in the future [6]. Although the calculation efficiency based on the physical model is high, it does not consider the driving style and behavior of the driver, and the physical model is only suitable for short-term prediction, such as vehicle danger warning. The behavior-based recognition model is a typical classifier. It uses the past position, motion state and context cues of the vehicle as features, and finally the trajectory prediction module outputs the future vehicle position and trajectory prediction [7-8]. A. Houenou et al. used heuristic classifiers for vehicle operation recognition, and used polynomial classifiers for trajectory prediction [9].C. Laugier et al. predicted vehicle trajectories based on Hidden Markov Model and Gaussian Mixture Model [10]. The behavior-based model is more accurate than the physical model, but the model does not consider the interaction between vehicles for modeling, which is not suitable for complex traffic environments [11]. Interaction-aware models take into account the impact of inter-vehicle interactions on vehicle motion to predict the vehicle's future trajectory. E. Kafer et al. implicitly learned from a large amount of traffic trajectory data by integrating the method of inter-vehicle interaction, and finally predicted the future trajectory of the vehicle [12].

Since motion prediction can be viewed as a sequence classification or sequence generation task, in recent years, many recurrent neural network (RNN) based methods have been proposed for maneuver classification and trajectory prediction. Alahi et al. proposed a social pooling LSTM network, which jointly models and predicts pedestrian movements in dense crowds by using social pooling layers [13]. N. Deo uses social convolutional pooling layers instead of social pooling The transformation layer is used to encode the historical trajectories of the surrounding vehicles and achieve accurate prediction of the future trajectories of the vehicles [14].

This paper introduces an attention mechanism based on the social convolutional pooling layer to capture the important information of the surrounding vehicle trajectory, and assigns different weights to the surrounding vehicle information, so that the network can adaptively extract the maximum trajectory features, thereby improving the prediction accuracy.

II. MODEL FRAME

Fig. 1 shows the attention-based social convolutional pooling network, which consists of three parts, which are composed of an LSTM encoder, a social convolutional pooling network with an attention mechanism, and a behavior-based LSTM decoding module.

Taking the historical trajectory information of the vehicle as input, the LSTM encoder uses the historical trajectory to learn the behavior of the vehicle and capture the motion state of the vehicle, while the social convolutional module aggregates the LSTM state of the target and the vehicles around the target into a social tensor to solve the problem of

978-1-6654-5087-4/22/$31.00 ©2022 IEEE

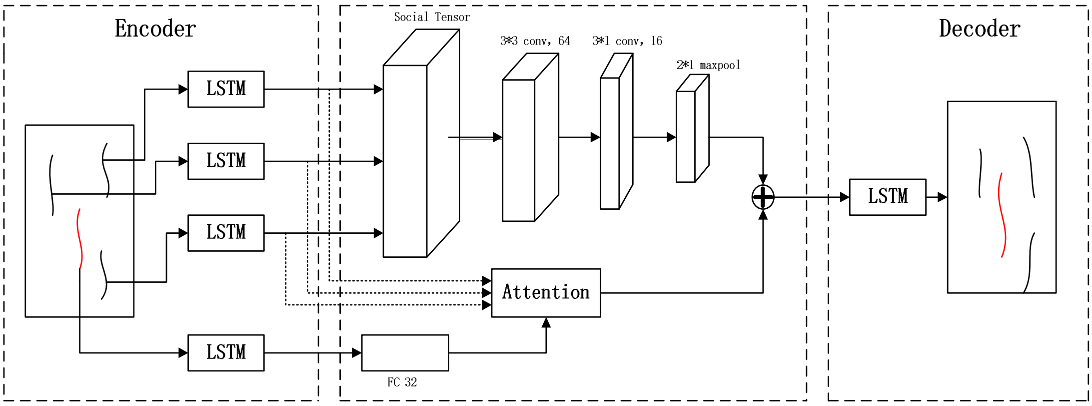

Fig. 1 attention-based social convolutional pooling network

the vehicle relationship between them. The attention mechanism in this paper judges the similarity of the historical hidden layer between the target vehicle and the surrounding vehicles, improves the local feature information of the social convolutional pooling network, and improves the surrounding vehicle information. The LSTM decoder predicts the following two types of vehicle behavior probabilities, namely lateral maneuvering and vertical maneuvering, as shown in Fig. 2. Lateral maneuvers are divided into left lane change, left acceleration lane change, right lane change, and right acceleration lane change; vertical maneuvering behavior is divided into braking and going straight; one-hot vector for lateral maneuvering and one-hot vector for vertical maneuvering are obtained, And concatenate them to calculate the final prediction result.

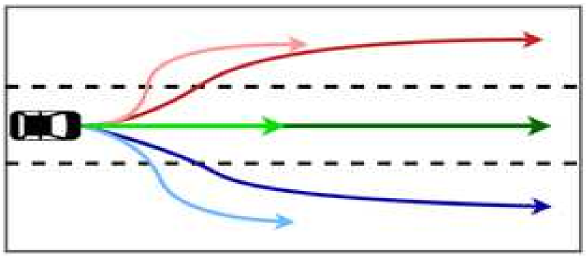

Fig. 2 Vehicle predictive trajectory maneuver classes

A. LSTM encoder module

We use the LSTM encoder to learn the characteristics of the vehicle movement. The clips of each frame of historical trajectories will predict the state of their vehicles and surrounding vehicles through the LSTM encoder. The LSTM status of each car is updated by frames in the past frame, and the LSTM network of each vehicle is used, and the share weight is shared. Encoder input is the historical trajectory characteristics of the vehicle.

$$X = \{ X_1, X_2, X_3, \ldots, X_n \} \quad (1)$$

X_i is the historical trajectory of the first vehicle.

$$X_i = \{x_i^t, y_i^t\} \quad (2)$$

Where x and y are the x and y coordinates of the vehicle at the current time t respectively.

The encoder consists of a LSTM network with the embedded module, and the historical track point (x, y) of the two-dimensional vehicle is embedded by the embedding layer to obtain a high dimensional space vector, and the high dimensional vector is delivered to the LSTM network to obtain a node The hidden layer, the formula is as follows:

$$e_t = FC(x_t, y_t, W_e) \quad (3)$$

$$h_t = LSTM(h_{t-1}, e_t, W_{em}) \quad (4)$$

The FC is the full connection layer mapping function in the embedded layer, the LSTM is the LSTM network mapping function, and $W_e$is the weight of the full connect layer in the encoder, and e_t is the high dimensional vector expression of the current time, h_{t-1} is The vector, W_{em} is the hidden layer weight of the LSTM encoder, h_t is a hidden vector for the current time. The characteristics obtained by the encoder are:

$$H = (h_1, h_2, \ldots h_{t-1}, h_t) \quad (5)$$

H is the historical hidden layer.

B. Attention social convolutional pooling module

The network model we use adds an attention model based on [14] to score important information about vehicle motion dynamics, and then assigns weights to the surrounding vehicle motion dynamics, weighted summation, and finally sends it to the social convolutional pooling module and extracts fusion of vehicle feature information.

Hidden layer of history of all vehicles around the target vehicle:

$$N = (H_1, H_2, \ldots H_{m-1}, H_m) \quad (6)$$

Attention network to obtain the motion dynamic correlation between vehicles.

$$S_i = f^a(H_i, H) \quad (7)$$

In the above formula, $H_i \epsilon N$ is the hidden layer of surrounding vehicle history. f^a is the calculation formula of the difference between the target vehicle and the surrounding vehicles, and the difference between the hidden layers is represented by the cosine distance between the vehicle and the surrounding vehicles, and normalized.

$$S_i = \cos(H_i, H) \quad (8)$$

$$\widetilde{S_i} = \frac{S_i}{\sum_{i=1}^{m} S_i} \quad (9)$$

The obtained normalization results are weighted and summed, and finally fused with the social convolutional pooling layer, the results are as follows:

$$e_{merge} = cs\_pool(h_1, h_2, \ldots, h_m) + \sum_{i=1}^{m} \widetilde{S_i}\, h \quad (10)$$

cs_pool is the social convolutional pooling network, $h_1, h_2, \ldots, h_m$ is the last hidden layer of the historical trajectory obtained by LSTM encoding.

C. decoder module

The LSTM-based decoder is used to generate a future motion prediction that predicts the next frame t, assuming that the prediction time is t=$T_{obs}+T$, and T_{obs} is the total number of observation time steps.

$$h_t = LSTM(h_{t-1}, e_{merge}, W_d) \tag{11}$$

The information of the hidden layer h_{t-1} at time t-1 is known, W_d is the weight of the LSTM decoder, h_t is the output of the LSTM network at the current prediction time, and h_t is finally mapped to obtain the predicted trajectory coordinates.

D. Probability Distribution Forecast

Assuming that y_t is the predicted trajectory coordinate mapped by h_t, the predicted trajectory collection Y is:

$$Y = \left(y_t, y_{t+1,}, \dots, y_{t+n}\right) \tag{12}$$

Calculate the probability distribution of the target vehicle trajectory according to the Gaussian distribution:

$$P_\theta(Y|m) = N(Y|\mu(X), \sum X) \tag{13}$$

$$\theta = (\theta^t, \theta^{t+1}, \dots, \theta^{t+n}) \tag{14}$$

m is the different driving behavior, and the driving behavior is divided into braking, acceleration, left lane change, right lane change, etc.; θ is the parameter of the binary Gaussian distribution of each time step in the future, corresponding to the mean and variance of the future position.

Conditional distribution probabilities for future trajectories of the model:

$$P(Y|X) = \sum P(m|X) P_\theta(Y|m, X) \tag{15}$$

E. The loss function of the model

Within the forecast horizon, the results of the forecast are shown as the root mean squared error (RMSE) of the forecast trajectory relative to the ground truth future trajectory. Since the LSTM model produces a binary Gaussian distribution, the mean squared error is calculated using the mean of the Gaussian components

$$RMSE = \frac{1}{n}\sum_t^{t+n} |(x_t^* - x_t)^2 + (y_t^* - y_t)^2| \tag{16}$$

(x_t^*, y_t^*) is the real trajectory coordinate, (x_t, y_t) is the predicted trajectory coordinate.

III. EXPERIMENTS AND RESULTS

A. Dataset

We use the publicly available NGSIM US-101 and I-80 datasets for experiments. The data set provides the world coordinates of the vehicle, each data set is 45 minutes in total, the sampling frequency is 10HZ, that is, 10 data points are reserved per second. Each dataset consists of 15 minutes of light, moderate and heavy traffic conditions. A training sample is 8s, with 3s (30 time steps) of historical trajectory and 5s (50 time steps) of predicted trajectory. Split the entire dataset into training, validation and test sets with a split ratio of 7:1:2.

B. Training details

We train the model using Adam mini-batch gradient descent with a learning rate of 0.001. In order to reduce the difficulty of training, during training, a data point is reserved every 0.2s, that is, 40 points are collected in 8s. The encoder LSTM has a 64-dimensional hidden state, and the decoder has a 128-dimensional hidden state. The size of the convolutional social pooling layer is shown in Figure 1. The size of the fully connected layer to obtain the vehicle dynamics encoding is 32. This experiment is trained on the GTX1080TI graphics card, and the algorithm model is implemented using the PyTorch framework.

C. Compared models

In order to verify the improvement effect of the method in this paper, this paper uses several methods of predicting the trajectory of the LSTM encoding and decoding structure as a comparison reference, which are listed as follows:

(1) LSTM: Use the classic encoding-decoding model framework to predict future trajectories

(2) S-LSTM: Represents Social LSTM, using social convolutional LSTM network to extract surrounding vehicle information for future trajectory prediction

(3) S-GAN: Represents Social GAN, which predicts trajectories by combining sequence prediction and generative adversarial networks.

(4) CS-LSTM: Represents Convolutional Social LSTM, social convolutional pooling method for trajectory prediction

(5) AS-LSTM: The improved method in this paper introduces the Attention structure in the convolutional pooling layer of Convolutional Social LSTM, which improves the way of extracting the original surrounding vehicle information

D. Model bias comparison

The experiment compares the trajectory mean square of each model in the 5s prediction time and compares the error value. The result is calculated every 1 second, 5 points are collected every second, and the average mean square error of 5 points is calculated every second, in meters. , the test results are shown in Table 1.

TABLE I shows the RMSE values for different models. The RMSE values of S-LSTM, S-GA, CS-LSTM and AS-LSTM are all better than the common LSTM network without social pooling structure, indicating the effectiveness of the proposed model; except for the LSTM network, other network models are It uses the motion information of the vehicle around the target vehicle, which shows that considering the

TABLE I. MODEL TRAJECTORY PREDICTION MEAN SQUARE ERROR COMPARISON

prediction algorithm	**1s**	**2s**	**3s**	**4s**	**5s**
LSTM	0.693	1.725	2.912	4.724	6.846
S-LSTM	0.622	1.613	2.777	4.069	5.974
S-GAN	0.583	1.549	2.582	3.967	5.376
CS-LSTM	0.476	1.257	2.147	3.426	4.751
AS-LSTM	0.339	0.991	1.771	2.750	3.993

interaction between vehicles is helpful for the prediction of future trajectories.

We can observe that the RMSE value of the AS-LSTM network is better than the RMSE value of the unimproved CS-LSTM network, indicating the importance of the attention mechanism introduced on the basis of CS-LSTM to obtain information about surrounding vehicles, increasing the global Vehicle characteristics, ultimately reducing the predicted trajectory measurement deviation.

IV. CONCLUSIONS

We propose a social convolutional pooling network based on attention mechanism in the vehicle trajectory prediction model of LSTM encoding and decoding based on social convolutional pooling network. By assigning different weights to the vehicles around the target vehicle and adding the feature information of all vehicles, the historical trajectory can be better used to learn the behavior of the vehicle, and finally the interdependence information of the spatial vehicle motion extracted by the social convolutional pooling layer is carried out. fusion. Comparing experiments with various models on two large public vehicle datasets, the results of the experiments are that the performance of the AS-LSTM model based on the attention mechanism in this paper is better than other models. However, the limitation of this paper is that it completely relies on the trajectory information of the vehicle to predict the future trajectory. Some visual information can be used to complement each other to improve the accuracy of predicting the future trajectory of the vehicle.

REFERENCES

[1] Y. Chai, B. Sapp, M. Bansal, and D. Anguelov, "Multipath: Multiple probabilistic anchor trajectory hypotheses for behavior prediction,"arXiv preprint arXiv:1910.05449, 2019.

[2] J. Nilsson, J. Silvlin, M. Brannstrom, E. Coelingh, and J. Fredriksson. If, when, and how to perform lane change maneuvers on highways. IEEE Intelligent Transportation Systems Magazine, 8(4):68–78, 2016. 1.

[3] N. Deo and M. M. Trivedi, "Convolutional social pooling for vehicle trajectory prediction," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2018.

[4] S. Lef`evre, D. Vasquez, and C. Laugier. A survey on motion prediction and risk assessment for intelligent vehicles. Robomech Journal, 1(1):1, 2014. 2.

[5] Bahram M, Hubmann C, Lawitzky A, et al. A combined model-and learning-based framework for interaction-aware maneuver prediction[J]. IEEE Transactions on Intelligent Transportation Systems, 2016, 17(6): 1538-1550.

[6] Batz, Watson, Beyerer. Recognition of dangerous situations within a cooperative group of vehicles[J]. Intelligent Vehicles Symposium IEEE, 2009: 907-912.

[7] Batz, Watson, Beyerer. Recognition of dangerous situations within a cooperative group of vehicles[J]. Intelligent Vehicles Symposium IEEE, 2009: 907-912.

[8] Q. Tran and J. Firl. Online maneuver recognition and multimodal trajectory prediction for intersection assistance using non-parametric regression. In Intelligent Vehicles Symposium Proceedings, 2014 IEEE, pages 918–923. IEEE, 2014.2

[9] A. Houenou, P. Bonnifait, V. Cherfaoui, andW. Yao. Vehicle trajectory prediction based on motion model and maneuver recognition. In Intelligent Robots and Systems (IROS), 2013 IEEE/RSJ International Conference on, pages 4363–4369.IEEE, 2013. 2

[10] C. Laugier, I. E. Paromtchik, M. Perrollaz, M. Yong, J.-D. Yoder, C. Tay, K. Mekhnacha, and A. N`egre. Probabilistic analysis of dynamic scenes and collision risks assessment to improve driving safety. IEEE Intelligent Transportation Systems Magazine, 3(4):4–19, 2011. 2.

[11] Klingelschmitt S, Platho M, Gross H-M, Willert V, Eggert J (2014) Combining behavior and situation information for reliably estimating multiple intentions In: Proc. IEEE Intelligent Vehicles Symposium, pp 388–393

[12] E. K¨afer, C. Hermes, C. W¨ohler, H. Ritter, and F. Kummert.Recognition of situation classes at road intersections. In Robotics and Automation (ICRA), 2010 IEEE International Conference on, pages 3960–3965. IEEE, 2010. 2

[13] A. Alahi, K. Goel, V. Ramanathan, A. Robicquet, L. Fei-Fei,and S. Savarese. Social lstm: Human trajectory prediction in crowded spaces. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 961–971,2016. 1, 2, 3, 6

[14] N. Deo and M. M. Trivedi, "Convolutional social pooling for vehicle trajectory prediction," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2018.