Attention-Based GRU for Driver Intention Recognition and Vehicle Trajectory Prediction

HAO Zixu
College of Communication Engineering
Jilin University
Changchun, China
haozx18@mails.jlu.edu.cn

Huang Xing
College of Communication Engineering
Jilin University
Changchun, China
xinghuang19@mails.jlu.edu.cn

WANG Kaige
College of Communication Engineering
Jilin University
Changchun, China
wangkg19@mails.jlu.edu.cn

CUI Maoyuan
Corporation-Intelligent Connected Vehicle Development Institute
China FAW Group
Changchun, China
cuimaoyuan@faw.com.cn

TIAN Yantao
College of Communication Engineering,
Key Laboratory of bionic engineering of Ministry of Education
Jilin University
Changchun, China
tianyt@ jlu.edu.cn

2020 4th CAA International Conference on Vehicular Control and Intelligence (CVCI) | 978-1-7281-8497-5/20/$31.00 ©2020 IEEE | DOI: 10.1109/CVCI51460.2020.9338510

***Abstract*—In human-machine cooperative decision making and control of intelligent vehicle, the intelligent system needs to understand driver's intention and desired vehicle trajectory in order to assist driver with safety driving in complex traffic scenes. In this paper, a vehicle trajectory prediction encoder-decoder model based on Gated Recurrent Unit (GRU) with attention mechanism is proposed. The proposed model is comprised of intention recognition module and trajectory prediction module. The intention recognition module was employed for recognizing driver's intention and calculating the probabilities of turning-left, lane-keeping, turning-right. The trajectory prediction module predicts vehicle trajectory using GRU decoder with attention mechanism, which takes vehicle historical position as input and predicts future position. Both intention recognition module and the trajectory prediction module share one encoder to save time. The NGSIM dataset was employed for training and testing. The experimental results indicate, comparing with traditional methods, the proposed horizontal-longitudinal decoupling hierarchical trajectory prediction method based on GRU neural network can predict driver's desired vehicle trajectory in a long prediction horizon and the attention mechanism improve the trajectory prediction accuracy at the same times.**

Keywords—trajectory prediction, intention recognition, Gated Recurrent Unit, attention mechanism

I. INTRODUCTION

In human-machine cooperative decision making and control of intelligent vehicle, the distribution of control rights between human and machine is a key issue. During the running of the intelligent vehicle, there will be a fight between driver and intelligent system for control rights when the intelligent system have disagreements with driver on decision, which is offensive to the original intention of invention of intelligent system because this reduce the safety of intelligent vehicle. Therefore, the key of decision making in human-machine cooperative control of intelligent vehicle is the intelligent system understands the intention of the driver [1]. As the supplement of the intention of the driver, the desired vehicle trajectory of the driver can assist the intelligent system makes better decision further. But, there are many traffic participants in traffic scenes where the interactions between these traffic participants are complex, which make the prediction of the driver's desired vehicle trajectory become a challenging issue.

Among the methods for trajectory prediction, it can be divided into two basic approaches, one is model-base method such as Kalman Filtering. Quadratic vehicle motion model [2] like Constant Turn Rate and Velocity model and Constant Turn Rate and Acceleration model can be solved by Bayesian Filtering [2], Extended Kalman Filter [3] or Unscented Kalman Filter [4] et.al, which the Kalman Filter can provide short-term vehicle motion prediction by calculating the prediction equation. But the disadvantages of model-base methods are obvious, these methods only provide vehicle future motion for one timestep, which is too short for decision making of the intelligent system.

Data-base methods is another way for prediction. Generative model like Mixtures of Gaussians [5] or Hidden Markov Model (HMM) [6] can be used for prediction. Artificial Neural Network and Deep-learning for sequence prediction like or LSTM [7] can also predict trajectory. Other methods including Inverse Reinforcement Learning [8] are based on explore mechanism. Some of data-base methods consider the uncertainty of future vehicle motion [5], which the uncertainty is brought by drivers or traffic scenes. Others are using complex information like 3D Radar point clouds [9] or images [8], which require expensive sensors and high-performance processors.

This paper proposes a vehicle trajectory prediction encoder-decoder model based on GRU with attention mechanism. The method only needs relative positions based on vehicle which come from High Precision Map such as GPS. It takes historical information as input and outputs vehicle predict position in future. In addition, both intention recognition module and the trajectory prediction module share one encoder to save time. The attention mechanism improves the prediction accuracy. Here are some reasons why not use other information. First, traffic scenes are complex, most information from other traffic participants require more sensors, sometimes these sensors cannot measure other participants 's information because of target occlusion, either.

Supported by the Joint Funds of the National Natural Science Foundation of China (Grant No. U19A2069 and Grant No. U1664263)
Corresponding author: TIAN Yantao.

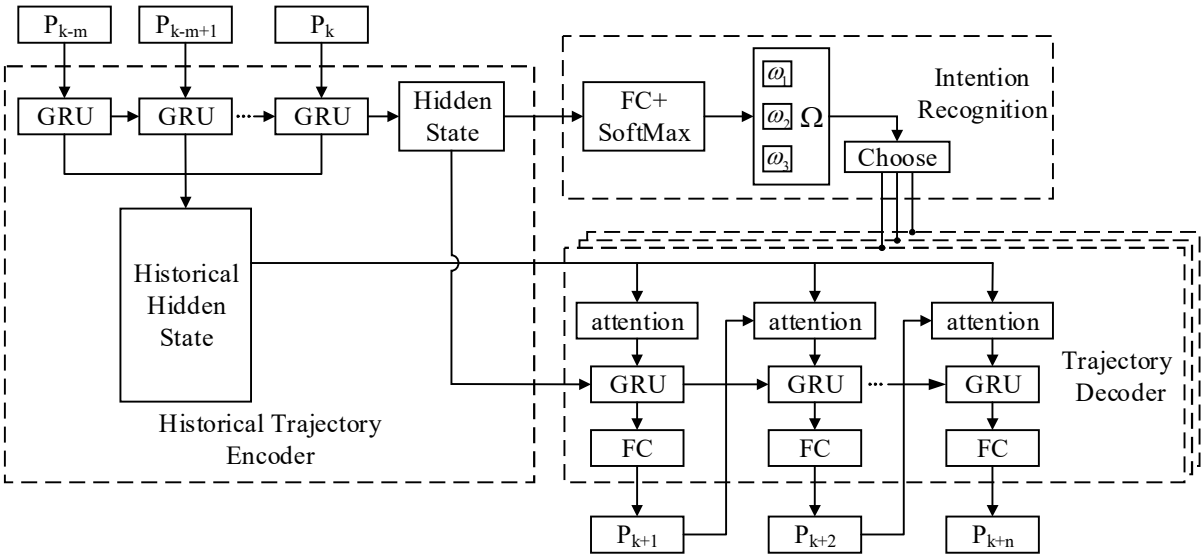


Fig. 1. The framework of the encoder-decoder model based on GRU with attention mechanism

The paper is structured as follow: Section II introduces the framework of the model. Section III discuss the method used in the model. Section IV contains the experiment and analysis of the model. The paper concludes in section V.

II. Framework

First of all, the driver's behaviors are defined as $C = \{LL, LK, LR\}$, which LL is *turning-left*, LR is *turning-right*, LK is *lane-keeping*. The historical information of vehicle position is given as

$$T_{k-m:k} = [P_{k-m}, P_{k-m+1}, \ldots, P_k] \tag{1}$$

where $P_k = [x_k, y_k]$ is the x and y co-ordinates at time k. The model's output is given as

$$T_{k:k+n} = [P_k, P_{k+1}, \ldots, P_{k+n}] \tag{2}$$

The general model this paper proposed show in Fig. 1. First, the historical trajectory encoder based on GRU takes historical trajectory $T_{k-m:k}$ as input. The encoder calculates the coding information and send to the intention recognition module, which combine by Fully Connect (FC) layer and SoftMax layer. The intention recognition module output the most probable intention and choose the corresponding trajectory prediction module. The trajectory prediction module takes the coding information and output the prediction trajectory.

III. Method

A. Historical trajectory encoder and intention recognition module

The intention recognition module is designed for learning how to recognize the driver's intention from data. Because the intention recognition module and the trajectory predict module need the same information of the historical trajectory i.e. $T_{k-m:k}$, they share the same encoder in order to save time.

The encoder is a Recurrent Neural Network (RNN) base on GRU [10]. At time t, the GRU take vehicle position P_t as input and output the hidden state $h_t^{encoder}$, which is given as

$$h_t^{encoder} = f_{GRU}(h_{t-1}^{encoder}, P_t) \tag{3}$$

GRU of the encoder takes P_t and $h_{t-1}^{encoder}$ as input and outputs $h_t^{encoder}$. The GRU controls the information flow from the previous activation when computing the new, candidate activation, but does not independently control the amount of the candidate activation being added [10]. This can decrease the computing time and keep the ability of reducing the difficulty due to vanishing gradients.

The output of GRU hidden state $h_t^{encoder}$ is the coding vector, which will be the input of intention recognition module and trajectory prediction module. The intention recognition module is combined by fully connect layer and SoftMax layer, which output the probability of different intention as intention probability vector and output the most possible intention of the driver.

The hidden unit of GRU in encoder is set to 128, the amount of the FC layer is set to 3 which have 128 neurons in each layer. The final layer is a linear layer with a SoftMax layer. The activation function is ReLU, and the loss function is set to Cross Entropy Error Function.

B. Trajectory prediction module

The trajectory prediction module needs to predict long term trajectory and outputs it in short time, so this paper use the decoder of the seq2seq model and the attention mechanism, which has widely used in NLP [11]. The decoder is also combined with GRU. In current time k, the model starts to predict, take the final hidden state of the encoder $h_k^{encoder}$ as initial hidden state of the decoder and current vehicle position P_k as input of the decoder. The GRU output the hidden state $h_k^{decoder}$, which is given by

$$h_k^{decoder} = f_{GRU}(h_k^{decoder}, P_k) \tag{4}$$

The FC layer takes the hidden state $h_k^{decoder}$ and predict next position of vehicle $\hat{P}_{k+1}$.Then, $\hat{P}_{k+1}$ and $h_k^{decoder}$ will be used to predict $\hat{P}_{k+2}$.

In theory, if computing time is enough, the decoder can predict vehicle position iteratively for a long time, which

means n can be infinity. But the recording of the historical information in RNN is attenuation, the decoder will forget more information of historical trajectory due to the predict time length increasing. At the same time, the prediction error is increasing.

The hidden unit of GRU in encoder is set to 128, the amount of the FC layer is set to 3 which have 128 neurons in each layer. he activation function is ReLU, and the final layer is a Liner layer.

C. Attention mechanism

As described in B, the hidden state will lose historical information when the predict time length increase. In order to reducing the difficulty of this, this paper introduce attention mechanism [12].

At time k, the historical hidden state of encoder is given by

$$H_{k-m:k}^{encoder} = [h_{k-m}^{encoder}, h_{k-m+1}^{encoder}, \dots, h_{k}^{encoder}] \quad (5)$$

The attention mechanism forces on correlativity between vehicle position P_t at time t and the historical information i.e. historical hidden state of the encoder $H_{k-m:k}^{encoder}$, which is given by

$$e_k^t = a(P_t, H_{k-m:\ k}^{encoder}) \quad (6)$$

where a is the correlation operator, in this paper, weighted dot-product is used, which convert (6) into

$$e_k^t = P_t \cdot W \cdot (H_{k-m:\ k}^{encoder})^T \quad (7)$$

Then, the SoftMax layer is used to calculate attention score $\overline{e_k^t}$ from e_k^t as follow:

$$\overline{e_k^t} = softmax(e_k^t) \quad (8)$$

The decoder with attention mechanism takes X_t as input, which X_t is the information combined with vehicle position P_t and the historical hidden state of the encoder $H_{k-m:\ k}^{encoder}$ with attention score $\overline{e_k^t}$

$$score_t = \overline{e_k^t} \cdot H_{k-m:\ k}^{encoder} \quad (9)$$

$$X_t = [P_t, score_t] \quad (10)$$

Now, at time t, the hidden state of decoder $h_t^{decoder}$ is given by

$$h_t^{decoder} = f_{GRU}(h_{t-1}^{decoder}, X_t) \quad (11)$$

IV. Experiment

This paper employs Interstate 80 Freeway Dataset in NGSIM project for training and testing. NGSIM stakeholder groups identified the collection of real-world vehicle trajectory data, which included vehicle trajectory, vehicle speed el at. The study area was approximately 500 meters (1,640 feet) in length and consisted of six freeway lanes, including a high-occupancy vehicle (HOV) lane [13], show in Fig .2.

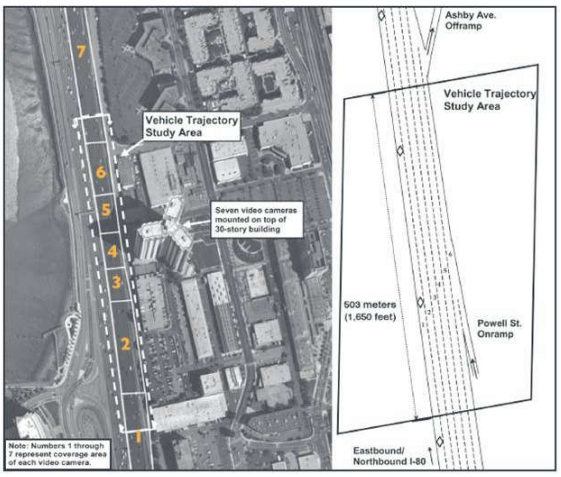


Fig. 2. Study area of I-80 high way [13]

A. Experiment pretreatment

The sampling period of dataset is 10hz, which the high frequency may weaken the prediction of the vehicle trajectory. First, the trajectory are resampled and the frequency are adjusted into 5hz. Then, *sEMA* [14] is employed as a data filter for trajectory data. Next the lane ID is extracted and the vehicle lane-changing points are found, i.e. the sampling position of vehicle which the lane ID of current host car is different as before.

Once the vehicle changing points are determined, the positions where the vehicle start to change the lane and where changing action end are determined by calculating the course angle of the vehicle θ, which is given by

$$\theta = \arctan\left(\frac{x_t - x_{t-3}}{y_t - y_{t-3}}\right) \quad (12)$$

If the course angle of the vehicle θ satisfy $|\theta| \le \theta_s$ for 3 times of continuous, the first point satisfying the condition will be the start point or end point of the changing process trajectory, which is defined by vehicle heading. The vehicle position points between the start point and the end point are defined as changing process points. What's more, some sample points need to be extracted before the changing process start point for evaluation of the intention recognition module in the early stage. According to this method, 764 changing processes of the vehicle are extracted, which *turning-left* for 117 and *turning-right* for 647. In order to achieve maximize the utilization of data, the data is slicing by sliding window which the length of window is n+m.

B. Analysis of intention recognition module

First of all, the performance of intention recognition module is compared with SVM and MLP. The evaluation indexes include accuracy rate, precision rate, recall rate and F1-score. The historical trajectory length is set to 1s. The confusion matrix of different method and different setup time for changing process is show in Table. 1. To evaluate the performance of different method in multiclass and class-imbalance issue, the evaluation indexes use the weighted macro form. The evaluation indexes of different method show in Table. 2.

TABLE I. THE CONFUSION MATRIX OF DIFFERENT METHOD

Setup time length (s)			Real behavior								
			Lane-keeping			*Turning-left*			*Turning-right*		
			1	1.5	3	1	1.5	3	1	1.5	3
Recognized Intention	*Lane-keeping*	GRU	2140	2230	2608	14	61	167	61	171	504
		SVM	2144	2234	2719	39	79	211	57	165	611
		MLP	2136	2244	2659	30	123	223	67	197	519
	Turning-left	GRU	11	21	40	260	284	350	3	2	5
		SVM	0	7	6	235	266	305	3	2	2
		MLP	7	1	4	245	222	286	4	2	1
	Turning-right	GRU	17	39	219	1	0	9	1696	1884	2327
		SVM	24	49	142	1	0	10	1700	1890	2223
		MLP	25	45	204	0	0	17	1689	1858	2316
Num of samples			2168	2290	2867	275	345	526	1760	2057	2836

TABLE II. THE EVALUATION INDEXES OF DIFFERENT METHOD

Method		Evaluation indexes											
		accuracy rate (%)			*precision rate (%)*			*recall rate (%)*			*F1-score (%)*		
		GRU	SVM	MLP	GRU	SVM	MLP	GRU	SVM	MLP	GRU	SVM	MLP
Setup Time (s)	1	97.43	97.05	96.81	97.44	97.10	96.81	97.43	97.05	96.81	97.42	97.03	96.81
	1.5	93.97	93.56	91.75	93.98	93.85	91.95	93.97	93.56	91.75	93.95	93.52	91.55
	3	84.54	84.24	82.95	85.07	86.19	83.73	84.54	84.24	82.95	84.33	84.04	82.78

TABLE III. THE EVALUATION INDEXES OF GRU IN DIFFERENT SETUP TIME

Setup time length (s)		Evaluation indexes								
		precision rate (%)			*recall rate (%)*			*F1-score (%)*		
		1	1.5	3	1	1.5	3	1	1.5	3
Real behavior	*Lane-keeping*	97.60	93.08	79.94	97.92	91.05	90.51	97.77	92.05	84.90
	Turning-left	95.47	87.62	89.57	92.00	82.03	63.69	93.70	84.73	74.44
	Turning-right	97.90	92.16	90.46	98.06	95.38	83.22	97.99	93.74	86.69

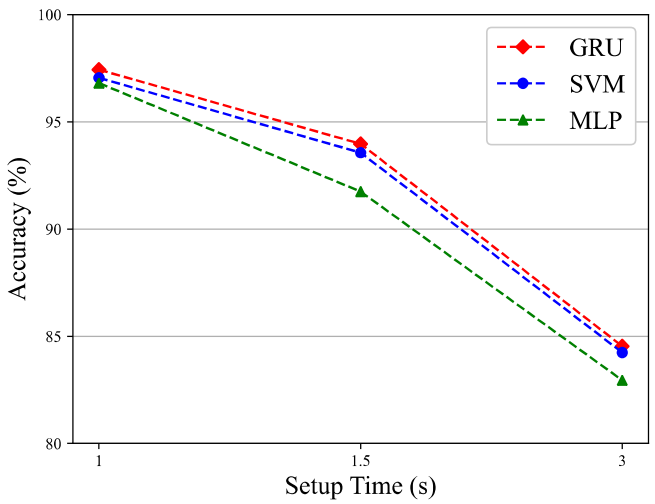


Fig. 3. Accuracy of different methods in different setup time

According to Table. 2 and Fig.3. The method we proposed have some advantage in these indexes than SVM and MLP. The setup time for changing process affect performance obviously, which the trajectory during the setup time is easy to be recognized as *lane-keeping*. The performance of the model we propose in different time and different class show in Table. 3.

Precision rate reflects the performance of the model in different class, and *recall rate* reflects the recognition rate of positive class. In different setup time, the recognition rate of *turning-left* is far less than *lane-keeping* and *turning-right*, because vehicles in I-80 dataset preferred to turn left i.e. change lane to fast lane 1 and 2.This make samples of *turning-left* less than *turning-right* and *lane-keeping*. When the setup time increased, the *precision rate* of *lane-keeping* decreases faster than the other, but *recall rate* doesn't decrease much. That means more samples are recognized as *lane-keeping*, because the features of the trajectory during the setup time are inconspicuous.

C. Analysis of trajectory prediction module

Root Mean Square Error (RMSE) is employed as the evaluation indexes between different methods. Compare the method in this paper with these methods

1) N-GRU: nomal RNN based on GRU

*2) N-ENDE:*encoder-decoder struct without intention recognition module and attention mechanism

*3) I-ENDE:*the framework in this paper without attention mechanism

*4) IA-ENDE:*the complete model in this paper including intention recognition module and attention mechanism

The setup time and the historical trajectory length are both set to 1s. The performance of different methods show in Table. 4 and Fig. 4.

N-GRU represents traditional model of prediction which provide vehicle future motion for one timestep. When comes to long term prediction, the error increase faster than others. N-ENDE predict trajectory in all direction, which weaken the performance of the model. I-ENDE can overcome the

disadvantage of N-GRU and N-ENDE, but show weakness while the setup time length is long. IA-ENDE can perform better than others because of the attention mechanism.

TABLE IV. RMSE OF DIFFERENT METHODS

Predict time length(s)	RMSE (m)			
	N-GRU	*N-ENDE*	*I-ENDE*	*IA-ENDE*
1	0.42	0.41	0.31	0.28
2	1.27	0.65	0.55	0.50
3	1.88	0.98	0.96	0.83
4	4.77	1.54	1.31	1.24

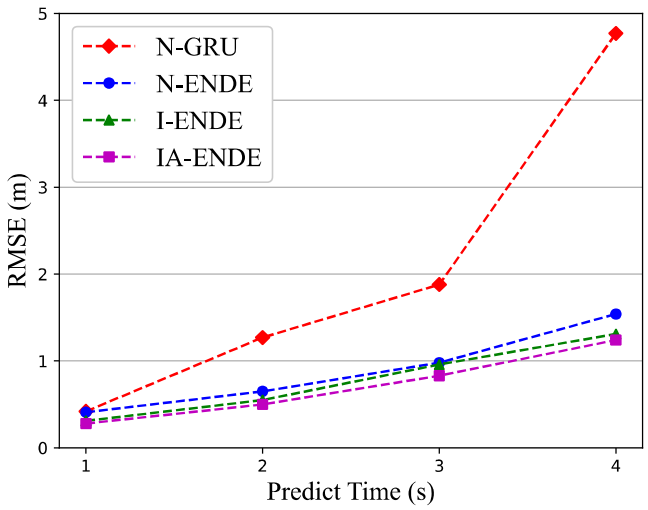


Fig. 4. The performance of different prediction methods in different predict time length

D. *Analysis of the complete model*

This paper selects a complete changing process trajectory for analysis, shown in Fig. 5.

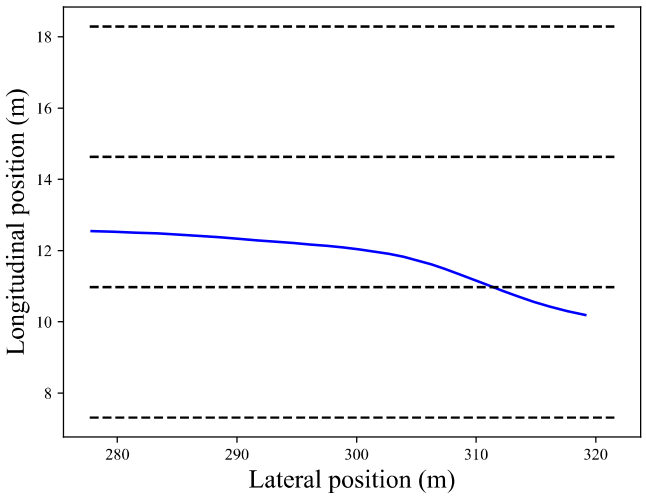


Fig. 5. The orginal trajectory

The setup time and the historical trajectory length are both set to 1s and predict trajectory for 2s. Once prediction finish, the historical window will slide to next sample point and predict trajectory. The result of the intention recognition show in Fig. 6.

The intention recognition module recognizes driver behavior at 1.8s, which 0.6s earlier before reaching the changing point and 0.8s later after changing start point, where the changing process is just in beginning. This indicates that the intention recognition module has low delay time.

Fig.7 shows the RMSE curve in whole prediction process. The max RMSE value is below 1.2m at time 0.8s to 1.6s i.e. the changing process is just ready to begin. At this time, the intention recognition is easy to recognize intention as lane keeping, as shown in Fig.8 (b).

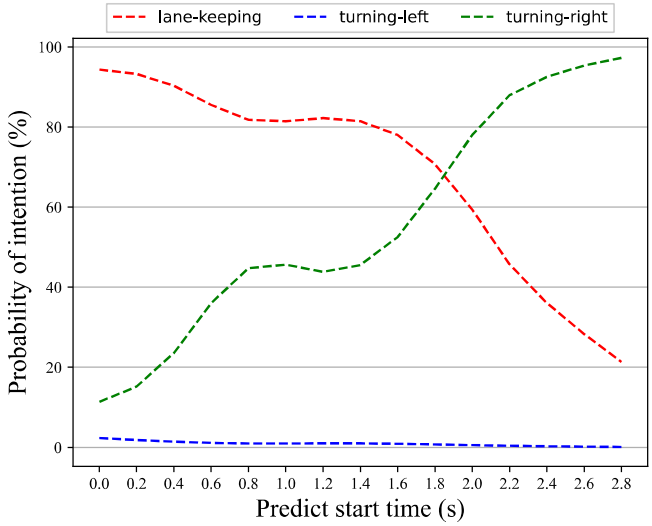


Fig. 6. Results of intention recognition

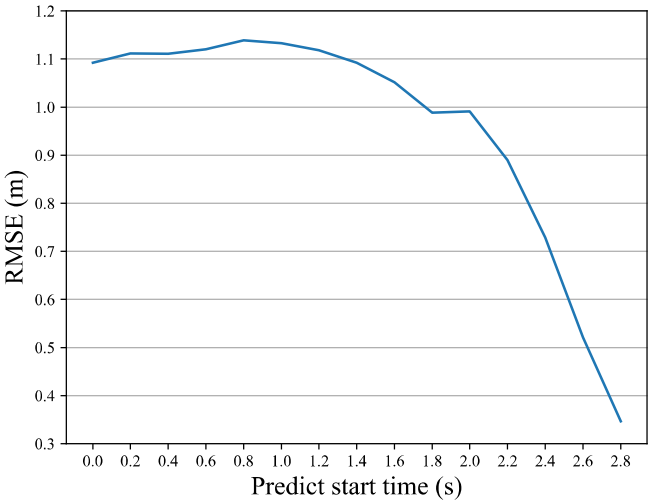
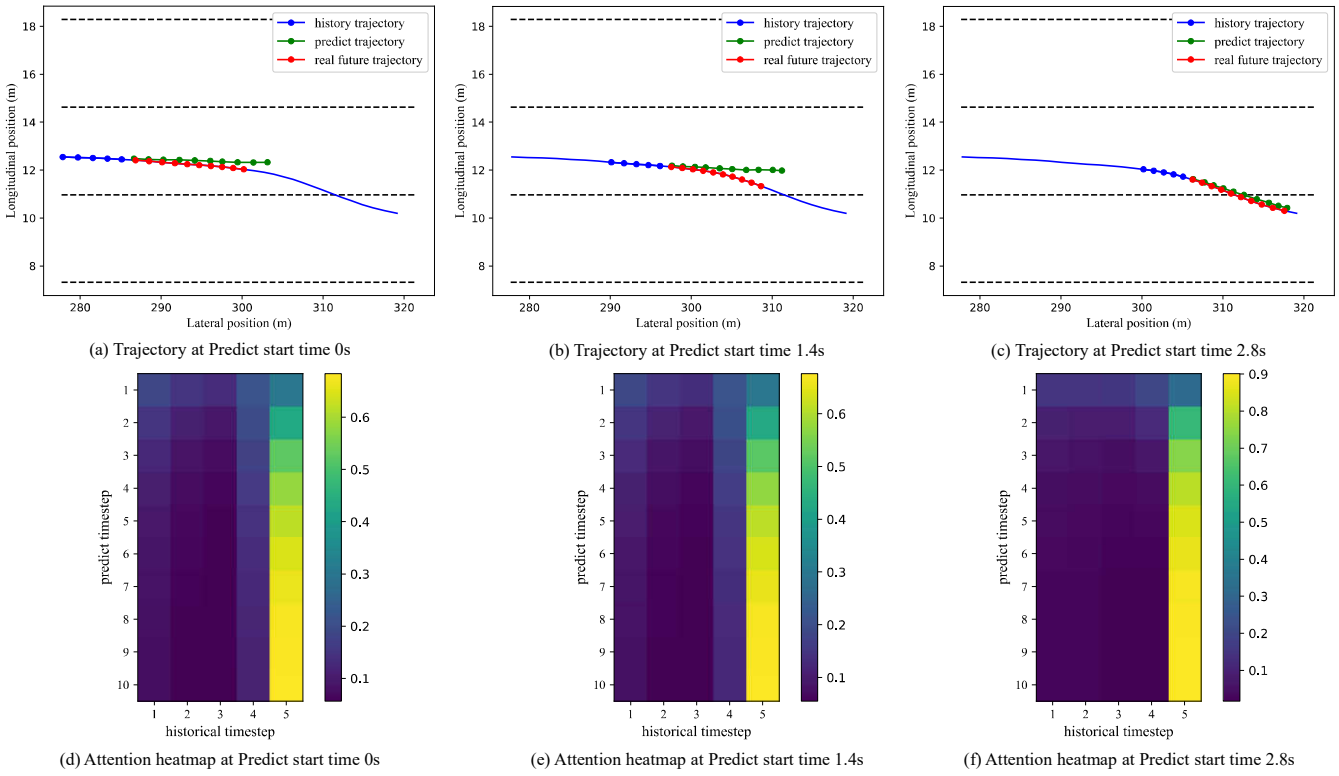


Fig. 7. RMSE curve in whole prediction process

Fig.8 show the particular time prediction and its attention visualization matrix, which attention visualization matrix display as heatmap. The visualization matrix reflects the correlation between the predict position and the historical trajectory. For later predict position, the model prefers to force on last historical trajectory position. This is similar to Markov Assumption. it is noteworthy that although the model force on latest information, the whole historical trajectory information still participates in prediction, rather than only the hidden state of decoder is taken as input, just like I-ENDE.



(a) Trajectory at Predict start time 0s

(b) Trajectory at Predict start time 1.4s

(c) Trajectory at Predict start time 2.8s

(d) Attention heatmap at Predict start time 0s

(e) Attention heatmap at Predict start time 1.4s

(f) Attention heatmap at Predict start time 2.8s

Fig. 8. Prediction results and heatmap of attention matrix

V. CONCLUSION

This paper introduces a vehicle trajectory prediction encoder-decoder model based on GRU with attention mechanism, which the intention recognition module and the trajectory prediction module share one encoder to reduce algorithm complexity. The attention mechanism brings more historical information for trajectory prediction module and improve the accuracy of prediction. Experiment show that the method in this paper has advantages in accuracy. In future work, we will discuss more about attention mechanism and bring it to interaction behavior prediction.

REFERENCES

[1] Y. HU, T. QU, J. LIU, Z. SHI, B. ZHU, D. CAO, H. CHEN. Human-machine Cooperative Control of Intelligent Vehicle: Recent Developments and Future Perspectives. Acta Automatica Sinica, 2019, 45(7), pp. 1261-1280.

[2] R. Schubert, C. Adam, M. Obst, N. Mattern, V. Leonhardt and G. Wanielik, "Empirical evaluation of vehicular models for ego motion estimation," 2011 IEEE Intelligent Vehicles Symposium (IV), Baden-Baden, 2011, pp. 534-539, doi: 10.1109/IVS.2011.5940526.

[3] K. Zindler, N. Geiß, K. Doll and S. Heinlein, "Real-time ego-motion estimation using Lidar and a vehicle model based Extended Kalman Filter," *17th International IEEE Conference on Intelligent Transportation Systems (ITSC)*, Qingdao, 2014, pp. 431-438, doi: 10.1109/ITSC.2014.6957728.

[4] H. Yu, L. Xie, J. Chen, C. Song and F. Guo, "Visual Odometry based on improved feature matching and Unscented Kalman Filter," *2016 35th Chinese Control Conference (CCC)*, Chengdu, 2016, pp. 5446-5450, doi: 10.1109/ChiCC.2016.7554203.

[5] J. Wiest, M. Höffken, U. Kreßel and K. Dietmayer, "Probabilistic trajectory prediction with Gaussian mixture models," 2012 IEEE Intelligent Vehicles Symposium, Alcala de Henares, 2012, pp. 141-146, doi: 10.1109/IVS.2012.6232277.

[6] B. Jiang and Y. Fei, "Traffic and vehicle speed prediction with neural network and Hidden Markov model in vehicular networks," 2015 IEEE Intelligent Vehicles Symposium (IV), Seoul, 2015, pp. 1082-1087, doi: 10.1109/IVS.2015.7225828.

[7] N. Deo, A. Rangesh and M. M. Trivedi, "How Would Surround Vehicles Move? A Unified Framework for Maneuver Classification and Motion Prediction," in IEEE Transactions on Intelligent Vehicles, vol. 3, no. 2, pp. 129-140, June 2018, doi: 10.1109/TIV.2018.2804159.

[8] L. Sun, W. Zhan and M. Tomizuka, "Probabilistic Prediction of Interactive Driving Behavior via Hierarchical Inverse Reinforcement Learning," 2018 21st International Conference on Intelligent Transportation Systems (ITSC), Maui, HI, 2018, pp. 2111-2117, doi: 10.1109/ITSC.2018.8569453.

[9] M. Bahram, A. Lawitzky, J. Friedrichs, M. Aeberhard and D. Wollherr, "A Game-Theoretic Approach to Replanning-Aware Interactive Scene Prediction and Planning," in IEEE Transactions on Vehicular Technology, vol. 65, no. 6, pp. 3981-3992, June 2016, doi: 10.1109/TVT.2015.2508009.

[10] K. Cho, B. Van Merri ̈enboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, "Learning phrase representations using rnn encoder-decoder for statistical machine translation," arXiv preprint arXiv:1406.1078, 2014.

[11] S. Gu and F. Lang, "A Chinese Text Corrector Based on Seq2Seq Model," 2017 International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery (CyberC), Nanjing, 2017, pp. 322-325, doi: 10.1109/CyberC.2017.82.

[12] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin. "Attention is all you need," 2017 Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, 4-9 December 2017, Long Beach, CA, USA, pp. 5998-6008.

[13] Federal Highway Adminstration, Interstate 80 Freeway Dataset, December 2006, Accessed on: June 13, 2020. [Online]. Available: https://www.fhwa.dot.gov/publications/research/operations/06137/index.cfm

[14] C. Thiemann, M. Treiber and A. Kesting, "Estimating acceleration and lane-changing dynamics based on ngsim trajectory data". Transportation Research Record Journal of the Transportation Research Board, 2088(2088), pp. 90-101.