Vehicle Trajectory Prediction Using Intention-based Conditional Variational Autoencoder

Conference Paper · October 2019

DOI: 10.1109/ITSC.2019.8917482

Xidong Feng
Tsinghua University

Vehicle Trajectory Prediction Using Intention-based Conditional Variational Autoencoder

Xidong Feng[1], Zhepeng Cen[1], Jianming Hu[1*] and Yi Zhang[1]

***Abstract*— Vehicle trajectory prediction has been an active research area in autonomous driving. In a real traffic scene, autonomous vehicle needs to predict future motion of surrounding vehicles before motion planning to improve driving safety and efficiency. In this paper, we first modify a sequence to sequence (seq-to-seq) maneuver-based model to produce possibility prediction of vehicle trajectory in future 5 seconds. Then we propose a novel method based on conditional variational autoencoder (CVAE). Our model generates multi-modal trajectory possibility prediction with high interpretability according to the estimation of driver's latent intention. Finally, we experiment the model on public traffic dataset and compare it with prior methods on trajectory prediction. The results show a great improvement on both lateral and longitudinal motion prediction, which also demonstrates the effectiveness of our model.**

I. INTRODUCTION

With the development of autonomous vehicles, accurately predicting surrounding vehicles' trajectory is becoming more and more important since this ability can instruct the autonomous vehicles to plan motion in advance after taking into consideration all possible behaviors of other vehicles. This technology benefits autonomous vehicle both in safety and in efficiency.

Previous work tries to utilize dynamic model to depict the physical nature of the trajectory prediction. However, this kind of methods do not account for the long-term trajectory because of the uncertainty of driver's intention. Thus, some data driven machine learning and deep learning methods are proposed to solve the long-term prediction problem based on the data gained from real-world vehicles' trajectories. However, given the same trajectory for previous few seconds, the trajectory in later few seconds can vary determined by many factors such as the interactions between vehicles, driver's habit, etc. Therefore, there is a growing trend in recent research that more probabilistic and multi-intention models are proposed. This type of models can readily help us with the uncertainty of prediction problem, which is also the foundation of our research.

This paper makes two contributions on the prediction problem. First, we modify the maneuver-based seq-to-seq model of previous research and eventually increase the performance on the same dataset. Then a new prediction method is proposed based on the combination of seq-to-seq and deep generative model called conditional variational autoencoder to generate trajectory given previous trajectory and maneuver.

The paper is organized as follows: Section II presents some related work about previous research on trajectory prediction; Section III provides our formulation towards trajectory prediction problem; The models' architectures of proposed methods are presented in Section IV; Section V presents the experiment on real-world dataset NGSIM and the results and evaluations of that; Section VI concludes the paper.

II. RELATED WORK

In general, work concerning trajectory prediction can be divided into two different parts, the model based methods focusing on the kinematic or dynamic model of vehicles and machine learning methods concerning the use of big data to analyze the hidden patterns of vehicles' trajectory. The model based methods take into consideration the limit of vehicle's physical model. For instance, [1] applies the velocity Kalman filter to deal with the uncertainty of time series prediction, which depends on the velocity model of vehicles. Model based methods make progress on short-term trajectory prediction. but quickly fail when focusing on long term prediction because of the exponential uncertainty growth with time going on.

At present, increasing amount of deep learning methods are proposed to predict long term vehicle trajectory. Long Short-Term Memory (LSTM) neural networks [2] and Gated Recurrent Units (GRU) [3], which are classical variants of Recurrent Neural Network (RNN), play a crucial role in tackling time series prediction. Preliminary work considers prediction problem as a regression [4] or a classification problem and makes some progress. But there are still two problems. First, traditional RNN method has a disadvantage that the input and output time-steps are fixed so that time-step could not be changed once the network architecture is determined. Second, due to the uncertainty of driver's intention, real future trajectories probably vary based on the maneuver even if the previous trajectories are the same, which means both regression method and classification method fail when tackling multi-modal trajectory and both of the methods can only learn the *average strategy* for given data.

In order to solve the first problem, inspired by the work [5] about natural language processing, work like [6] begins to incorporate seq-to-seq model with trajectory prediction task, which allows input and output with unfixed time-steps. This

[1]X. Feng, Z. Cen, J. Hu (corresponding author) and Y. Zhang are with the Department of Automation, Tsinghua University, Beijing 100084, China. Email: (fxd16, cenzp16)@mails.tsinghua.edu.cn, (hujm, zhyi)@mail.tsinghua.edu.cn)

[2]J. Hu (corresponding author) and Y. Zhang are with the Beijing National Research Center for Information Science and Technology, Beijing, 100084, China.

architecture also improves the performance compared with previous vanilla LSTM work.

As for the second problem, more work begins to focus on the combination of driver intention or maneuver recognition and vehicle prediction trajectory. [7] uses a probabilistic graphical model combining both historical data and vehicle velocity to estimate intentions in ramp merging. [8] applies LSTM network to recognize potential lane-change intention using surrounding vehicle trajectories. [9] relies on a dual LSTM network to recognize the driver's intention and calculate lateral deviation of current vehicle and target lane accordingly as another input of prediction model. However, their models still regard the problem as a regression one so they still cannot deal with multi-modal problem. Therefore, some probabilistic trajectory prediction models are proposed to solve the second problem. [10] produces bivariate Gaussian distribution about velocity based on mixture density network, which can generate sample points of velocity and successfully presents diverse probability of future trajectory. Like proposed work in this paper, [11] combines maneuver probability prediction and conditional probability together to generate multi-modal trajectory with different probability density. But model architecture proposed in this paper makes the conditional probability module more interpretable and achieves better results.

Some other work like [12] utilizes conditional variational autoencoder [13], which is also the fundamental architecture of our model, to generate fixed coordinates of trajectory. But our work focuses on a different kind of task. Moreover, our model incorporates the predicted maneuver of driver and outputs a probability distribution to make trajectory prediction more interpretable.

III. PROBLEM FORMULATION

A. *Prediction problem*

The goal of vehicle trajectory prediction is to estimate the probability distribution of the future vehicle (ego vehicle in Fig. 1) position accordingly given previously observed trajectories and relative positions of ego and surrounding vehicles. Therefore, this prediction problem aims to learn a conditional possibility distribution $f(\cdot)$ of future outputs Y in T_{fut} time-steps given T_{hist} historical observation X

$$P(Y^{(t)}|\mathbf{X}), P(Y^{(t+1)}|\mathbf{X}), \cdots, P(Y^{(t+T_{fut}-1)}|\mathbf{X})$$

where

$$\mathbf{X} = \left[X^{(t+T_{hist}-1)}, \cdots, X^{(t-1)}\right]$$

B. *Inputs and outputs*

The inputs have T_{hist} time-steps observations, and each observation includes displacement, velocity and longitudinal distant Δx between ego and surrounding vehicles,

$$X^{(t)} = \left[x_0^{(t)}, y_0^{(t)}, vx_0^{(t)}, vy_0^{(t)}, \Delta x_0^{(t)}, x_1^{(t)}, \cdots, \Delta x_5^{(t)}\right]^T$$

Unlike other prediction model concerning surrounding vehicles such as [4] [8] [9], our model inputs observations

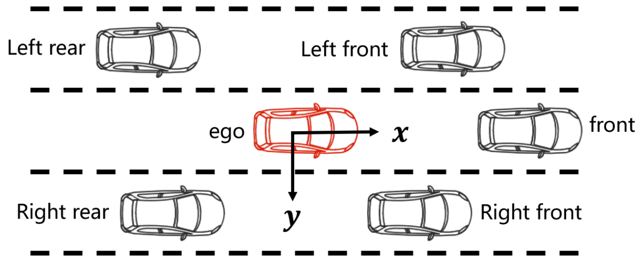


Fig. 1: Relative position of vehicles

of 6 vehicles shown in Fig. 1 instead. According to traffic regulations, the front vehicle takes no responsibility for the rear. Therefore, the trajectory or driving decision of the front one is usually influenced little by the rear one.

Considering longitudinal velocity is usually steadier than displacement, we choose longitudinal velocity and lateral displacement as our model outputs Y.

$$Y^{(t)} = \left[vx_0^{(t)}, y_0^{(t)}\right]^T$$

In our model, the lateral displacement of center line of current lane is set as 0 and we simply use the backward difference of displacement to calculate velocity,

$$vx^{(t)} = \nabla x^{(t)} = x^{(t)} - x^{(t-1)} \tag{1}$$

The inputs sequences T_{hist} and outputs sequences T_{fut} are both 25 in this model. But as mentioned before, the time-steps are not fixed because of the seq-to-seq model.

IV. PROPOSED METHOD

In this paper, we employ a LSTM network to recognize driver's latent intention at the beginning and modify the seq-to-seq model based on [11] for trajectory prediction. The modified model exceeds the original one but can only generate uni-modal trajectory. To deal with it, we propose a more interpretable multi-modal prediction model based on CVAE which even outperforms the modified seq-to-seq model on the estimation of trajectory probability distribution.

A. *Intention recognition network*

This network concentrates on recognition of driver's latent intention on lane-change, which is a classification problem. Our model divides the intentions into three classes including

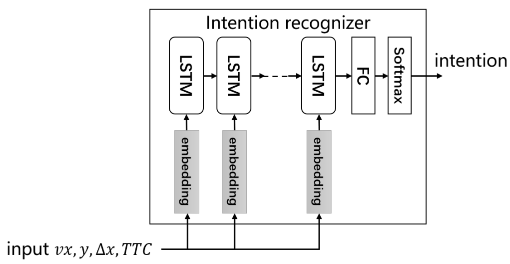


Fig. 2: intention recognizer

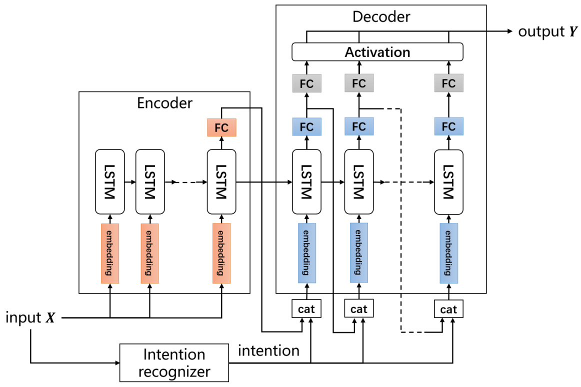


Fig. 3: **Modified seq-to-seq prediction model**: FC is a fully connected layer and cat is a concatenate layer concatenating two tensors. The final states of encoder will be replicated to decoder as initial states and meantime the decoder takes the final hidden states of encoder after a fully connected layer as the first input.

left and right lane-change and lane-keeping. If vehicle is in different lanes at the last moment of the input and real output, we define intention as lane-change. Similar to other classification problems, this network is trained by minimizing cross entropy between estimation and true intentions.

We use a LSTM network adopted in [8] as intention recognizer and present it in Fig. 2. But our inputs are quite different such as longitudinal velocity, lateral displacement of ego and surrounding vehicles and reciprocal of time-to-collision (TTC) at the last input instant, which is calculated by longitudinal distance and ego and surrounding vehicles velocity

$$TTC^{-1} = \frac{x_0^{t-1} - x_i^{t-1}}{\Delta x_i^{t-1}}, i = 1, 2, \cdots, 5 \tag{2}$$

Meanwhile, our inputs remove vehicle acceleration and steering angle, which can suggest intentions effectively in short term but performs poorly on longer horizons. Additional embedding layers is applied before LSTM.

B. *Modified maneuver-based seq-to-seq model*

This model with seq-to-seq framework is modified based on previous work [9] [11] and is presented in Fig. 3. Intention recognizer estimates latent driving intention and outputs a vector, and then decoder LSTM is updated by both intentions and the LSTM states at the previous instant. Therefore, the estimated intentions are inputted at each time-step to instruct the model to produce vehicle trajectory prediction. Compared with some seq-to-seq models such as [11] who repeat the hidden states of encoder as inputs to decoder in each time-step, our model has an advantage that last few outputs are influenced less by the hidden states of encoder which sometimes fail to predict potential uncertainty even if intention has been estimated. For example, if a vehicle driver keeps lane during the entire input time and intention recognizer also estimates intention as lane keeping, their models will probably fail to produce the possibility of lane-change in the last few time-steps of output. On the contrary, our seq-to-seq model performs more flexibly in long term.

At the end of model, we use Gaussian mixture model (GMM) to describe the possibility distribution of (vx, y)

$$P(vx, y) = \sum_{i=1}^{n} \alpha_i \phi_i(vx, y) \tag{3}$$

consisting of n bivariate Gaussian distributions

$$\phi_i(vx, y) = \mathcal{N}(\mu_{vx}, \mu_y, \sigma_{vx}, \sigma_{vx}, \rho) \tag{4}$$

The α_i notates the weight of the i-th Gaussian distribution. Therefore, an activation layer is needed to fulfill the constraints of possibility distribution

$$\alpha = \frac{\exp(h_i^{(\alpha)})}{\sum \exp(h_i^{(\alpha)})}, \mu = h^{(\mu)}, \sigma = \exp h^{(\sigma)}, \rho = \tanh h^{(\rho)} \tag{5}$$

where $h^{(\alpha)}, h^{(\mu)}, h^{(\sigma)}, h^{(\rho)}$ notate states before activation layer. This model adopts negative log-likelihood as loss function.

C. *Intention CVAE*

Modified seq-to-seq model performs better than other models in previous research. However, there still exists a problem that the output probability distribution cannot be distinguished by single maneuver even though the GMM can theoretically generate the combination of probability distribution of all maneuvers. From the experiment, we find due to the fact that the estimated variance for a single maneuver is so large that aliasing of different trajectory distributions from different maneuvers happens, which means we cannot tell the exact maneuver's trajectory distribution from GMM.

Therefore, we choose to turn to utilize deep generative model called conditional variational autoencoder (CVAE) [13]. The CVAE model is a variant of variational autoencoder (VAE) model, which introduces a new stochastic latent variable to represent the features of the input with lower dimension. Additionally, CVAE introduces the label (condition) of the input so the model can generate conditional outputs accordingly, which offers a more interpretable approach to produce multi-modal prediction than previous work.

For more details about VAE and CVAE, you can see [13]. Here the difference between the probability model we use

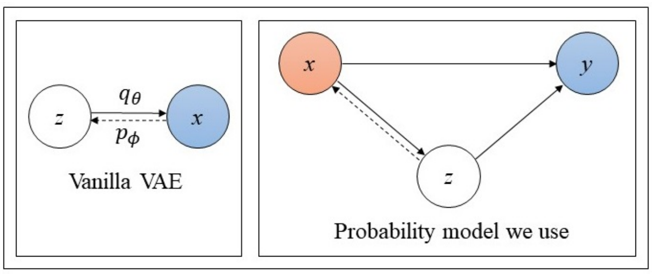


Fig. 4: Difference of probability model

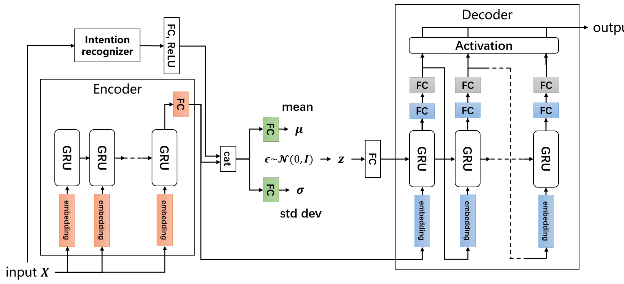


Fig. 5: Proposed intention CVAE

and vanilla VAE is shown in Fig. 4. We have to mention that vanilla generative model cannot produce prediction result. Vanilla VAE focus on approximates the posterior distribution of latent variable z by x and utilize z to regenerate x. Inspired by its structure, we leverage the probabilistic model shown on the right side of Fig. 4. We assume a latent variable z influences past trajectory x. Meanwhile, z and past trajectory influence future trajectory y. So the whole model we propose aims at first approximating the posterior distribution of z and then predicting future trajectory y based on z and past trajectory x. We also denote this structure as CVAE since they have exactly the same architecture.

The detailed model architecture is shown in Fig. 5. The main structure of CVAE is derived from the seq-to-seq model though LSMT is replaced with GRU, the latter of which is believed to converge faster than vanilla LSTM. Moreover, in order to highlight the importance of latent variable z, we modify the CVAE seq-to-seq model and transform the latent variable z into the initial hidden state of the decoder by a fully connected layer. Encoder's final hidden state is directly fed into the first input of the decoder to make the best use of the encoded information. The detailed structure of the decoder and encoder is the same as the modified seq-to-seq model in section B. The loss function is shown as follows.

$$\begin{aligned} l_i(\theta, \phi) = &- W_1 \mathbb{E}_{z \sim q_\theta(z|x_i, c)} \left[\log p_\phi \left(y_i | z, c\right)\right] \\ &+ W_2 \mathbb{KL} \left(q_\theta \left(z | x_i, c\right) \| p(z, c)\right) \end{aligned} \tag{6}$$

where θ, ϕ notate the parameters of the encoder and decoder and W_1, W_2 notate the coefficients of loss. The former part in loss function represents loss for likelihood and the input latent z for decoder comes from the encoder given the condition and previous time series data. The latter one represents the KL divergence between the multivariate normal distribution and the output distribution from the encoder because latent variable z is assumed to be subject to normal distribution. Coefficients W_1 and W_2 are manually set so that two loss components can have similar order of magnitude,

Our model distinguishes different maneuvers' trajectory by feeding different conditions into the model such as $[1, 0, 0]$ for turning left, $[0, 1, 0]$ for keeping the current lane and $[0, 0, 1]$ for turning right. At last, we incorporate the intention module into our CVAE model. The CVAE model can give out the trajectory distribution of different maneuvers, while the intention module will assign the probability of each. Based on that the sampling points can be generated.

V. EXPERIMENT

A. Dataset and Preparation

Next Generation Simulation (NGSIM) [14] dataset is one of the largest available sources of naturalistic vehicle trajectories, which includes 4 sub-datasets US-101, I-80, Lankershim and Peachtree. Each one has three sections and includes 45-minute trajectories data totally. We select first two for our experiment: we sample data from the whole set of US-101 and the last two sections of I-20 as training set while using the first section of I-20 as test set.

- **Time horizon and down-sampling**
  Every sample lasts 10s and we divide the first 5s track as model input and the last 5s as prediction horizon. Data in NGSIM are sampled of 10Hz, but we down-sample with a scale factor of 2 like [11]. Besides to simplify model, down-sampling can also reduce the noise of vehicle velocity effectively which is calculated by numerical difference. Therefore, our input horizon T_{hist} and output horizon T_{fut} are both 25.
- **longitudinal distance truncation**
  Because far distant vehicles usually have less influence on predicted trajectories, our model sets a maximum longitudinal distance d_{max} as truncation

$$\Delta x = \begin{cases} -d_{max}, \Delta x < -d_{max} \\ \Delta x, |\Delta x| \leq d_{max} \\ d_{max}, \Delta x > d_{max} \end{cases} \tag{7}$$

  where Δx notates longitudinal distance between ego vehicle and surrounding ones.
  Moreover, if there is no vehicle surrounding predicted one, a *dummy vehicle* will be placed at the maximum distance. For example, when there is no vehicle on the left lane, we place dummy vehicles at both left rear and left front. The maximum distance is set as 200ft (61m).
- **Data balance**
  Less than 3% of trajectories sampled completely randomly involve lane-change, which means the dataset has a serious problem of data imbalance. Therefore, we first randomly sample only lane-change trajectories and get a sampled-set, and then randomly sample all maneuvers of data as another. Finally, we combine these two sets (we use a ratio of 50%:50%) as our training set. As a result, the amount of lane-change trajectories is roughly equal to lane-keeping.

B. Compared Models

In the experiment, we compare four different models in previous research with two models in this paper as follows.

- **Constant velocity based model of Kalman filter**
  In this part, we implement the Kalman filter method and set the model as constant velocity.
- **Vanilla LSTM based model proposed in [4]**
  The vanilla LSTM model considers prediction as a regression problem and outputs all of the position for 5 seconds at the last time-step of a single LSTM.

- **Intention based LSTM model proposed in [9]**
  In this Model, the authors also take into consideration the intention of vehicles' driver. However, they still utilize the vanilla LSTM instead of seq-to-seq model.
- **Multi-Modal intention based seq-to-seq model proposed in [11]**
  Similar to ours, this work divides the situation according to intention and generates trajectory respectively. The intention module produces the probability for each intention and assigns it to each generated trajectory.
- **Modified intention based seq-to-seq model**
  We make several adjustments on original intention based seq-to-seq model and eventually get a better result on RMSE than previous research and consider this as a new baseline for our CVAE model.
- **Intention based CVAE**
  Besides the combination of intention prediction module and CVAE (CVAE(P)) where the future maneuver comes from the intention module, we also consider pure CVAE model with ground truth (CVAE(GT)) where the future maneuver comes from future true trajectory to show the performances of both CVAE model and the whole one.

Due to the fact that the latter three methods' output probability distribution, we adopt the calculation method in [11] that first chooses the maneuver with highest probability generated by the prediction module, then calculates the mean of the distribution and calculates the RMSE at last. Because of uncertainty in reproduction of neural network model, we just reproduce the model in [9] and directly adopt the results reported in the rest papers.

C. Result and Analysis

In this part, a few tables and figures are presented to illustrate the result of our experiment. The Root Mean Squared Error (RMSE) of longitudinal and lateral positions are used for comparative analysis among models.

Table. I shows the longitudinal position RMSE of $1 \sim 5$ seconds for models being compared. The result shows vanilla LSTM in fact has the worst performance in compared models, which means vanilla LSTM cannot model well for trajectory data with different maneuvers. The improvement of multi-modal module and our modified seq-to-seq model indicates the effects of auxiliary information from intention and considering the prediction separately based on maneuvers. Our intention-based CVAE makes progress over previous research and achieves the best result. Although CVAE(GT) uses the real maneuver from future data and should be excluded to comparison, it also shows the excellence of CVAE module in our model.

TABLE I: Longitudinal position error(m)

Method \ Time	1s	2s	3s	4s	5s
Kalman Filter	0.73	1.78	3.13	4.78	6.68
Vanilla LSTM in [4]	0.71	1.98	3.75	5.96	9.00
Intention LSTM in [9]	**0.58**	1.39	2.57	4.04	5.77
Multi-Modal in [11]	**0.58**	1.26	2.12	3.24	4.66
Modified Intention Seq2seq	0.65	1.39	2.25	3.32	4.62
Intention CVAE(P)	0.64	**1.21**	**2.01**	**3.07**	**4.33**
Intention CVAE(GT)	0.54	1.17	1.97	3.06	4.03

TABLE II: Lateral position error(m)

Methods	1s	2s	3s	4s	5s
Modified intention seq2seq	0.17	0.35	0.57	0.80	0.97
Intention CVAE(P)	**0.17**	**0.34**	**0.54**	**0.74**	**0.89**
Intention CVAE(GT)	0.14	0.22	0.28	0.34	0.40

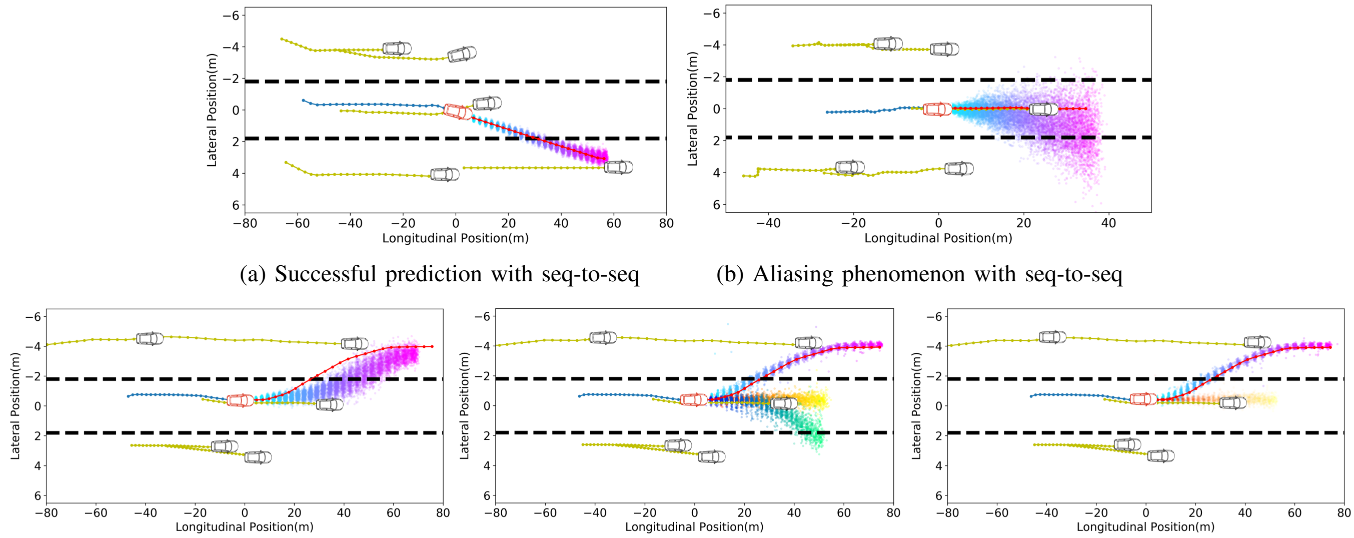
TABLE III: Lateral position error for different maneuvers(M)

Maneuver	Method	1s	2s	3s	4s	5s
Left	[9]	0.56	1.05	1.51	1.98	2.38
	M-Seq2seq	0.38	0.58	0.80	0.99	1.13
	CVAE(P)	**0.30**	**0.53**	**0.74**	**0.88**	**1.01**
Straight	[9]	**0.13**	**0.21**	**0.26**	**0.30**	**0.33**
	M-Seq2seq	0.16	0.34	0.56	0.78	0.95
	CVAE(P)	0.16	0.33	0.53	0.73	0.89
Right	[9]	0.51	1.14	1.73	2.43	2.96
	M-Seq2seq	0.39	0.52	0.68	0.78	1.02
	CVAE(P)	**0.17**	**0.34**	**0.57**	**0.76**	**0.95**

Table. II shows the lateral position RMSE for $1 \sim 5$ seconds for models proposed in this paper separately. The reasons for presenting the results alone is shown in Table. III, which shows lateral position error for different maneuvers. After reproducing models in previous researches which significantly outperform ours, we notice they neglected the problem of data imbalance. So their modules pay much attention to lane-keep maneuver and regard most of maneuvers as that, which seems to provide a better performance in the accuracy of intention module and lower RMSE for all of the trajectories because of the low proportion of lane-change maneuver. However, when dealing with lane-change maneuver, one of the most significant parts in trajectory prediction, the models will eventually perform much worse than the average error. Table. III evidently show this idea. Though [9] achieves an outstanding performance in lane-keep maneuver, its results in lane-change are much worse than ours which solves the problem of imbalanced data.

Although modified seq-to-seq and intention-based CVAE have similar performances in RMSE, the CVAE model still has advantages over the former which cannot be evaluated with single metric. Fig. 6 presents the generated probability distribution for modified seq-to-seq model and intention CVAE model. Fig.6a, 6b show the different kinds of result of modified seq-to-seq model. As mentioned in section IV, though modified seq-to-seq model can achieve prediction with low variance in Fig.6a, it still suffers from the aliasing problem in Fig.6b due to the uncertainty in future. Fig. 6d depicts all trajectory distributions given three maneuvers and Fig. 6e is the result after assigning maneuver's probability generated by intention module to each trajectory's distribution. We also provide Fig. 6c generated from modified seq-to-seq model to make comparison with intention CVAE on the same data, indicating intention CVAE can produce more accurate trajectory distribution without aliasing problem.

What's more, we believe by incorporating the scenarios' information like the width of road into conditions of CVAE



(a) Successful prediction with seq-to-seq (b) Aliasing phenomenon with seq-to-seq

(c) Modified seq-to-seq (d) All distribution with pure CVAE (e) Final distribution with intention CVAE

Fig. 6: **Results for modified Seq2seq and intention CVAE**: Surrounding vehicles and ego vehicle are denoted with symbols of different color. Blue lines and red lines denote the previous and future trajectory of ego vehicle separately and yellow lines denote previous trajectories of surrounding vehicles. Moreover, the gradient color represents different time-steps and we use three gradient colors to distinguish diverse maneuver's prediction in 6d,6e.

and enhancing the data, our algorithm can generalize and adapt to different scenarios.

VI. CONCLUSIONS

In this paper, we propose a maneuver-based prediction method based on CVAE, which naturally generates multi-modal vehicle trajectories according to driver's intention. We compare our method with previous work by experimenting on public highway dataset, and the results show our method outperforms the state of art in prediction error. Additionally, our method produces multi-modal prediction corresponding to uncertain intention of driver with higher accuracy and interpretability. Finally, we give specific analysis on results and case study in certain scene.

Future work towards trajectory prediction includes extending model to more complex traffic scenes such as intersections or mixed scene of pedestrian and vehicles, co-prediction of both ego vehicle and surrounding vehicles as a representation of local traffic condition and other prediction evaluation metrics.

VII. ACKNOWLEDGMENT

This work is partially supported by National Key R&D Program in China (2018YFB1600802, 2018YFB1600601), National Natural Science Foundation of China (61673232) and Tsinghua University Initiative Scientific Research Program (2018Z05JDX005, 20183080016).

REFERENCES

[1] S. Ammoun and F. Nashashibi, "Real time trajectory prediction for collision risk estimation between vehicles," in *2009 IEEE 5th International Conference on Intelligent Computer Communication and Processing*, pp. 417–422, IEEE, 2009.

[2] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[3] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio, "Empirical evaluation of gated recurrent neural networks on sequence modeling," *arXiv preprint arXiv:1412.3555*, 2014.

[4] F. Altché and A. de La Fortelle, "An lstm network for highway trajectory prediction," in *2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC)*, pp. 353–359, IEEE, 2017.

[5] I. Sutskever, O. Vinyals, and Q. V. Le, "Sequence to sequence learning with neural networks," in *Advances in neural information processing systems*, pp. 3104–3112, 2014.

[6] S. H. Park, B. Kim, C. M. Kang, C. C. Chung, and J. W. Choi, "Sequence-to-sequence prediction of vehicle trajectory via lstm encoder-decoder architecture," in *2018 IEEE Intelligent Vehicles Symposium (IV)*, pp. 1672–1678, IEEE, 2018.

[7] C. Dong, J. M. Dolan, and B. Litkouhi, "Intention estimation for ramp merging control in autonomous driving," in *2017 IEEE Intelligent Vehicles Symposium (IV)*, pp. 1584–1589, June 2017.

[8] S. Su, K. Muelling, J. Dolan, P. Palanisamy, and P. Mudalige, "Learning vehicle surrounding-aware lane-changing behavior from observed trajectories," in *2018 IEEE Intelligent Vehicles Symposium (IV)*, pp. 1412–1417, June 2018.

[9] L. Xin, P. Wang, C.-Y. Chan, J. Chen, S. E. Li, and B. Cheng, "Intention-aware long horizon trajectory prediction of surrounding vehicles using dual lstm networks," in *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*, pp. 1441–1446, IEEE, 2018.

[10] Y. Hu, W. Zhan, and M. Tomizuka, "Probabilistic prediction of vehicle semantic intention and motion," in *2018 IEEE Intelligent Vehicles Symposium (IV)*, pp. 307–313, IEEE, 2018.

[11] N. Deo and M. M. Trivedi, "Multi-modal trajectory prediction of surrounding vehicles with maneuver based lstms," in *2018 IEEE Intelligent Vehicles Symposium (IV)*, pp. 1179–1184, IEEE, 2018.

[12] N. Lee, W. Choi, P. Vernaza, C. B. Choy, P. H. Torr, and M. Chandraker, "Desire: Distant future prediction in dynamic scenes with interacting agents," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 336–345, 2017.

[13] K. Sohn, H. Lee, and X. Yan, "Learning structured output representation using deep conditional generative models," in *Advances in neural information processing systems*, pp. 3483–3491, 2015.

[14] V. Alexiadis, J. Colyar, J. Halkias, R. Hranac, and G. McHale, "The next generation simulation program," *ITE Journal (Institute of Transportation Engineers)*, vol. 74, pp. 22–26, 08 2004.