Article
http://justc.ustc.edu.cn
Received: June 30, 2021; Accepted: May 25, 2022

The surrounding vehicles behavior prediction for intelligent vehicles based on Att-BiLSTM

Yunqing Gao[1], Juping Zhu[1], and Hongbo Gao[1, 2] ✉

[1]*Department of Automation, University of Science and Technology of China, Hefei 230022, China;*
[2]*Institute of Advanced Technology, University of Science and Technology of China, Hefei 230088, China*

✉Correspondence: Hongbo Gao, E-mail: ghb48@ustc.edu.cn
© 2022 The Author(s). This is an open access article under the CC BY-NC-ND 4.0 license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Graphical abstract

Flowchart of the presented Att-BiLSTM algorithm.

Public summary

- BiLSTM model combined with encoder is adopted to ensure the memory of long-time series training.
- The importance of different states problem can be solved by considering the attention mechanism.
- The Att-BiLSTM model is designed to ensure the surrounding vehicles prediction effectiveness and accuracy.

Citation: Gao Y Q, Zhu J P, Gao H B. The surrounding vehicles behavior prediction for intelligent vehicles based on Att-BiLSTM. *JUSTC*, **2022**, 52(9): 6. DOI: 10.52396/JUSTC-2021-0155

JUSTC

Article

http://justc.ustc.edu.cn

Received: June 30, 2021; Accepted: May 25, 2022

The surrounding vehicles behavior prediction for intelligent vehicles based on Att-BiLSTM

Yunqing Gao[1], Juping Zhu[1], and Hongbo Gao[1, 2] ✉

[1]*Department of Automation, University of Science and Technology of China, Hefei 230022, China;*
[2]*Institute of Advanced Technology, University of Science and Technology of China, Hefei 230088, China*

✉Correspondence: Hongbo Gao, E-mail: ghb48@ustc.edu.cn
© 2022 The Author(s). This is an open access article under the CC BY-NC-ND 4.0 license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Cite This: *JUSTC*, **2022**, 52(9): 6 (8pp)

Read Online

Abstract: A surrounding vehicles behavior prediction method was presented for intelligent vehicles. The surrounding vehicles' behavior is hard to predict since the significant uncertainty of vehicle driving and environmental changes. This method adopts bidirectional long short-term memory (BiLSTM) model combined with an encoder to ensure the memory of long-time series training. By constructing an attention mechanism based on BiLSTM, we consider the importance of different information which could guarantee the encoder's memory under long sequence. The designed attention-bidirectional LSTM (Att-BiLSTM) model is adopted to ensure the surrounding vehicles' prediction accuracy and effectiveness.

Keywords: behavior prediction; attention mechanism; long short-term memory (LSTM); intelligent vehicle

CLC number: TP183 **Document code:** A

1 Introduction

Intelligent driving is one of the vital application fields of artificial intelligence (AI), which improves the speed of traffic, relieves traffic pressure, and significantly improves human life[1]. To ensure a safe driving path, it is necessary to predict the future driving behavior of surrounding vehicles in advance.

Scholars have done many works to achieve accurate and fast behavior prediction. Classical methods include three types: physical model, intention model, and interaction model[2]. The physical model is only based on the motion model of vehicles, without considering the interaction with other vehicles on the road and the restrictions of traffic rules[3]. Traffic rules are considered in the intention model, but there is no interaction with other motor vehicles. Moreover, it lacks qualitative constraints on time, so it can't adjust to roads with other topological structures. Predicting with the intention model may cause intelligent driving cars to plan a very unreasonable trajectory or collide with other motor vehicles[4]. Given the above shortcomings, the interaction model considers the interaction between traffic rules and other road participants, which takes longer to predict than the physical model, and is more stable than the prediction based on intention. The only disadvantage is that the calculation of the relationship between multiple cars is relatively large[5]. Fortunately, with the popularity of artificial intelligence and the rocket-like development of deep learning in recent years, intelligent vehicle behavior prediction has a more perfect and powerful tool. Because of good performance in complex and realistic scenes, these deep learning methods have become more important.

Deep neural networks have been applied to many machine learning tasks and have successfully learned various practical situation representations[6, 7]. The recurrent neural network (RNN) is widely used to analyze time series data structures. LSTM, as a popular RNN variant, has shown excellent performance in various tasks in vehicle behavior prediction[8]. Moreover, BiLSTM is a combination of forward LSTM and backward LSTM, which has been widely researched and used[9].

Facing existing problems, we present an innovative method that adds the attention mechanism to the encoder-decoder of bidirectional LSTM and makes different weights to different states' importance in the network training process. And we solve the problem that the encoder will forget if the sequence is too long. The overview of the developed algorithm is shown in Fig. 1. The historical sequence dataset is an input into the BiLSTM algorithm model, and the data training is realized in the feedback process based on the loss function. Then, based on the first 10 time steps, predict the 11th step, and so on, to get the prediction result.

BiLSTM and Att-BiLSTM are existing theoretical methods; we have applied this knowledge to the platform of intelligent driving, which made some innovations based on the original research. Based on the experimental results and analysis of surrounding vehicles' behaviors prediction, Att-BiLSTM theory has good performance in the field of intelligent driving. In addition, we have added encoder and decoder on the basis of Att-BiLSTM, which is also a great innovation.

The main contributions of this study are summarized as follows:

(ⅰ) BiLSTM model combined with encoder is adopted to ensure the memory of long-time series training.

(ⅱ) According to that, add the attention mechanism. Therefore, it solved the importance of different states' problems.

DOI: 10.52396/JUSTC-2021-0155
JUSTC, **2022**, 52(9): 6

Fig. 1. Att-BiLSTM overview.

(iii) The Att-BiLSTM model is designed, which ensures the effectiveness and accuracy of surrounding vehicles' prediction.

The structure of this work is as follows: Section 2 reviews the related work. Section 3 provides a detail procedure about the Att-BiLSTM model based on LSTM and the attention mechanism. Experiment results are shown in Section 4. Finally, Section 5 concludes this work and proposes the future work.

2 Related work

For LSTM, Ondruska and Posner[10] utilized this with ranging sensor measurement as input to track the position of an object, and Khosroshahi et al.[11] used LSTM with trajectory data as input to identify the driver's intention. However, they have not solved the problem that the model input is a fixed time step.

Jeong et al.[12] proposed an RNN algorithm based on LSTM to predict the motion of vehicles around multi-lane turning intersections. Ref. [13] designed a vehicle trajectory algorithm based on deep learning, where the encoder based on LSTM is used to analyze the past trajectory patterns, and the decoder based on LSTM is used to generate future trajectory sequences. The structure uses beam search technology to generate K most possible trajectory candidates on the occupied grid map so that K local optimal candidates are far away from the decoder output. Refs. [14, 15] presented a long-term prediction method based on a recursive neural network model of the gating unit. The deep neural network with LSTM and GRU structure is used to analyze the temporal and spatial characteristics of the past trajectory. The system that generates the future trajectory of other traffic participants through sequence learning can be used in different prediction levels. However, LSTM model often has the problem of gradient disappearance, so it isn't easy to train for a long time series. In Ref. [16], a trajectory prediction model based on spatio-temporal LSTM (ST-LSTM) was designed, which embeds spatial interaction into the LSTM model to implicitly measure the interaction between adjacent vehicles. Moreover, a quick connection between the input and output of two consecutive LSTM layers is introduced to deal with gradient disappearance. An algorithm for long-term trajectory prediction of surrounding vehicles was designed by using double LSTM network[17], which effectively improved the prediction accuracy in a strong interactive driving environment. Compared with the traditional trajectory matching and artificial feature selection methods, this method can automatically learn the advanced spatio-temporal features of driver behavior from natural driving data through sequence learning. Refs. [18, 19] proposed a coupled LSTM model using the historical track of the target vehicle and the surrounding environment information effectively to predict the future motion track of the vehicle. This method predicts the observable motion intention of the vehicle, builds a grid for the target vehicle, and extracts the hidden space and intention information of adjacent vehicles. However, they did not consider that the information of each time period in the input sequence has different contributions to the final prediction results. For example, in an input sequence, the predicted target vehicle has tried to change lanes at an earlier time and tended to follow later. So assigning different weights to the final prediction results for each time period is necessary. Therefore, the surrounding vehicles' behavior prediction method is presented for intelligent vehicles based on Attention-BiLSTM.

3 Behavior prediction model based LSTM and attention mechanism

In this section, We first introduce the RNN algorithm, then the LSTM algorithm and attention mechanism, and finally the Att-BiLSTM algorithm.

3.1 RNN algorithm

RNN is a cyclic neural network, which performs the same function on each data input, and the output depends on the

DOI: 10.52396/JUSTC-2021-0155
JUSTC, **2022**, 52(9): 6

last calculation. The generated output is copied and sent back to the circular network. In other neural networks, all inputs are independent of each other. But in RNN, all inputs are interrelated. In many real-life and work tasks, the network's output is related to the current input and previous output.

Given an input sequence $x_{1:T} = (x_1, x_2, \ldots, x_t, \ldots, x_T)$, RNN updates the activity value h_t of hidden layer with feedback edge by the following formula:

$$h_t = f(h_{t-1}, x_t), \tag{1}$$

where $h_0 = 0$, $f(^*)$ is a nonlinear function and can be a feed-forward network. Fig. 2 shows an example of RNN, where the "delayer" is a virtual unit that records the latest (or several) activity values of neurons.

3.2 LSTM network

The explicit design of LSTM is to solve the defect that the circulating neural network cannot deal with long-range problems. Fig. 3 shows the cyclic unit structure of LSTM network which introduces gating mechanism to determine the direction and path of information transmission. Denote the input gate as i_t, the forgetting gate as f_t, and the output gate as o_t.

The gate in the LSTM network is a kind of "soft" gate, and its value is between 0 and 1, which means that information passes through in a certain proportion. The calculation method of the three gates is shown in Eq. (2).

Fig. 2. Recurrent neural network.

Fig. 3. The cyclic unit structure of LSTM network.

$$\begin{aligned} i_t &= \sigma(W_i x_t + U_i h_{t-1} + b_i), \\ f_t &= \sigma(W_f x_t + U_f h_{t-1} + b_f), \\ o_t &= \sigma(W_o x_t + U_o h_{t-1} + b_o). \end{aligned} \tag{2}$$

The input gate i_t controls the amount of candidate state $\tilde{c}_t$ information at the current time. The forgetting gate f_t controls the internal state c_{t-1} at the last moment, selectively deletes some information in the storage unit, and activates it to the next input memory. Output gate o_t controls how much information the internal state c_t needs to output to the external state h_t at the current moment, as shown in Eq. (3).

$$\begin{aligned} c_t &= f_t c_{t-1} + i_t \widetilde{c}_t, \\ h_t &= o_t \tanh(c_t). \end{aligned} \tag{3}$$

LSTM network is an improved form of RNN, which makes the past data in the network easier to remember. LSTM solves the gradient disappearance problem in RNN. Using back-propagation to train the model makes LSTM suitable for classifying, processing, and predicting time series with unknown time delays.

3.3 Attention mechanism

The amount of information stored in neural networks is called network capacity. Generally speaking, the capacity of storing data with a group of neurons is proportional to the number of neurons and the complexity of the network. If more information is to be stored, the number of neurons will be more or the network will be more complex, leading to the multiple increase of neural network parameters.

When using a neural network to process a large amount of input information, we can learn from the attention mechanism of the human brain and select only some key information to process to improve the efficiency of the neural network. $X = [x_1, \ldots, x_N]$ presents N pieces of input information. To save computing resources, it is not necessary to input all N input information into the neural network for calculation, but only to select some task-related information from X and input it into the neural network.

The calculation of attention mechanism can be divided into two steps: One is to calculate the attention distribution on all input information, and the other is to calculate the weighted average of input information according to the attention distribution.

Let Key = Value = X, then attention distribution can be obtained as follows:

$$\alpha_i = \mathrm{softmax}(s(\mathrm{Key}_i, q)) = \mathrm{softmax}(s(X_i, q)), \tag{4}$$

where α_i is an attention distribution (probability distribution), $s(X_i, q)$ is the attention scoring mechanism. There are several scoring mechanisms:

Additive model: $s(X_i, q) = v^{\mathrm{T}} \tanh(WX_i + Uq)$.

Dot product model: $s(X_i, q) = X_i^{\mathrm{T}} q$.

Scale dot product model: $s(X_i, q) = \dfrac{X_i^{\mathrm{T}} q}{\sqrt{d}}$.

Bilinear model: $s(X_i, q) = X_i^{\mathrm{T}} W q$.

The attention distribution α_i can be interpreted as the degree to the i-th information is concerned when a task-related query q is given. We use a "soft" information selection mechanism to summarize the input information. The mechanism of

DOI: 10.52396/JUSTC-2021-0155
JUSTC, **2022**, 52(9): 6

soft attention is as follows:

$$\mathrm{att}(q,X)=\sum_{i=1}^{N}\alpha_i X_i. \tag{5}$$

In addition, some other changed models, for example, hard attention, key-value attention, etc., will not be described here. The attention mechanism is generally used as a neural network component, mainly to filter information and select relevant information from input information.

3.4 Att-BiLSTM with encoder-decoder structure

BiLSTM consists of two layers of long-term and short-term memory networks, which have the same input and different ways of information transmission. The encoder-decoder structure is adopted to ensure the memory of long-time series training. The encoder is composed of BiLSTM, base on the input X_t at time t, the output vector h_t is obtained through BiLSTM encoder. The calculation is shown in the following Eq. (6), where $\overrightarrow{h_t}$ represents the forward output, $\overleftarrow{h_t}$ represents the reverse output, and $\oplus$ represents the addition of corresponding elements.

$$h_t=\left[\overrightarrow{h_t}\oplus\overleftarrow{h_t}\right]. \tag{6}$$

In the structure of encoding and decoding, as the length of the input sequence becomes longer, the information of the input sequence saved by the intermediate vector obtained by the encoder will become less and less. At this time, by adding the attention mechanism, the decoder can no longer only consider the hidden state of the last time step of the encoder but can choose the most relevant hidden state among the hidden states of the input sequence at all times. Correspondingly, in this work, the model will use the hidden vectors of all time steps of the encoder at every moment of decoding and focus on the most relevant input information of the time step according to attention. The encoder output is added to the attention mechanism, assuming that $H=[h_1,h_2,...,h_t]$ is the set of BiLSTM outputs after t time steps. The calculation method of attention shows as follow:

$$\begin{aligned} M&=\tanh(H),\\ \alpha&=\mathrm{softmax}\left(w^{\mathrm{T}}M\right),\\ \tau&=H\alpha^{\mathrm{T}},\\ h^*&=\tanh(\tau). \end{aligned} \tag{7}$$

Input h^* to the decoder, then the prediction result is obtained. The whole process is shown in Fig. 4.

4 Experiment and result analysis

4.1 Datasets

In our experiment, the proposed model is tested on a public dataset: next generation simulation (NGSIM) which is widely based on many existing studies[20−22]. The NGSIM dataset is the detailed vehicle trajectory data collected by the Federal Highway Administration of the United States on the American highway 101 in Los Angeles, California. NGSIM contains four data sets in different scenes, namely, US101, I-80, Lankershim Boulevard, and Peachtree Street. The first two data sets record the driving track of vehicles on expressways; in comparison, the last two data sets record the driving track of vehicles on urban roads. And US-101 is selected as the primary research dataset, and the operations of other data sets are similar.

The vehicle trajectory feature information in the NGSIM dataset includes the target vehicle's global and local coordinates. Also, it contains essential information such as vehicle speed, acceleration, front and rear following distance, and so on. The study area consists of five main lanes and two auxiliary lanes. The vehicle trajectory data provides the precise position of each vehicle in the study area every tenth of a second, thus obtaining the precise lane position and the position relative to other vehicles.

4.2 Datasets preprocessing

Firstly, based on Python's Pandas tool, the US-101 dataset in the NGSIM dataset is extracted, and then the training set and test set of this experiment are constructed. The specific process is as follows:

(i) Arrange all vehicle numbers from small to large.

(ii) Arrange the total number of frames of the vehicle from small to large, and every vehicle entering the recording area will be numbered and reused, so we only take the data with the smallest total number of frames in each vehicle number and get 3108 sets of data.

(iii) Each data's total number of frames is different. The position data is recorded once every tenth second, so the data with existing frames less than 60 are discarded, the first 60 data with more than 60 are taken, and finally, 2100 groups are obtained.

(iv) Take 2000 groups of data, and take random training from them. The remaining 100 groups are used as test set data.

4.3 Model training

Based on the obtained training set processing, the final train-

Fig. 4. Attention mechanism-based BiLSTM.

DOI: 10.52396/JUSTC-2021-0155
JUSTC, **2022**, 52(9): 6

ing sample contains historical information of 10-time steps and future trajectory information of 1-time step. The specific process is shown in Algorithm 4.1.

Algorithm 4.1: Generation of training samples.

Data: InputData *tw*
Result: InoutSeq
InoutSeq=[]
L = len(InputData)
for *i* in range(0, $L-tw-1$) **do**
    trainSeq = InputData[i: $i+tw$]
    trainLabel = InputData[$i+tw$: $i+tw+1$]
    InoutSeq.append((trainSeq, trainLabel))
end

For Algorithm 4.1, InputData presents the input data, and *tw* presents the sequence length of the input encoder, i.e., 10. The dimension of the hidden state of encoder and decoder is 128. Adam optimizer is used in training, and the learning rate is 0.001. The algorithm is based on PyTorch. The model's task is to predict the trajectory of intelligent driving vehicles. Naturally, the loss function of the model is set as the mean square error between the predicted future trajectory and the true value of the trajectory, as shown in Eq. (8):

$$\text{Loss} = \frac{\sum_{i=1}^{k}\left((x_{\text{pre}}-x_{\text{Tr}})^2+(y_{\text{pre}}-y_{\text{Tr}})^2\right)}{k}, \tag{8}$$

where $(x_{\text{pre}}, y_{\text{pre}})$ is the predicted trajectory coordinate, $(x_{\text{Tr}}, y_{\text{Tr}})$ is the true value of the trajectory.

The losses of the final model in the training set and test set are shown in Fig. 5. It shows that the loss of the model in the training set and test set decrease continuously with the increasing of iteration times and finally tend to be stable. This process indicates that the verification result can be completed fast after the model training.

4.4 Experiment results and analysis

We compared the proposed Att-BiLSTM and the classical LSTM prediction methods based on the matplotlib tool provided by python. For a specific car number (3000 or 2000), the prediction results are shown in Figs. 6 and 7, where Figs. 6a and 7a are the prediction result obtained based on Att-BiLSTM, and Figs. 6b and 7b are the result obtained by LSTM method. The blue point in the image is the actual route, and the red point is the predicted route. Prediction and prediction_1 indicate that the data of the previous 10 time steps are input to predict the data of the next time step, and every 10 time steps are predicted. prediction_2 is to take the data of 10 time steps at the beginning and predict the data of 10 time steps in the future, which is only conducted once.

(a) Training set loss change curve. (b) Test set loss change curve.

Fig. 5. Loss change curve of training and test sets.

(a) The presented behavior prediction model prediction result. (b) LSTM algorithm prediction result.

Fig. 6. The prediction results with ID 30000 car.

DOI: 10.52396/JUSTC-2021-0155
JUSTC, **2022**, 52(9): 6

(a) The presented behavior prediction model prediction result. (b) LSTM algorithm prediction result.

Fig. 7. The prediction results with ID 20000 car.

Table 1. The results of calculating the average distance error.

Evaluation index	Prediction time step	LSTM	Att-BiLSTM
ADE	1	23.072	12.343
	2	36.764	19.699
	5	107.384	58.163

They show that Att-BiLSTM is indeed much better than the LSTM prediction method. The algorithm can well predict the future trajectory of a simple cycle track and predict the complex cycle track with multiple direction changes. Although the predicted trajectory obtained by the algorithm has some deviation in position, the overall direction is still accurate.

To reflect the superiority of this algorithm as a whole, we compare the following contents with the classic LSTM prediction algorithm in the test set from the data level. Average displacement error (ADE) will be used as an index to evaluate the prediction results in the experiment, which is defined as the Euclidean distance between the predicted track coordinates of the predicted vehicle in the future time step and the real value, i.e.,

$$\mathrm{ADE} = \frac{\sum_{j=1}^{N}\left(\sum_{i=1}^{k}\left((x_{\mathrm{pre}}-x_{\mathrm{Tr}})^2+(y_{\mathrm{pre}}-y_{\mathrm{Tr}})^2\right)\right)}{N\times k}, \tag{9}$$

where N represents the number of samples, $(x_{\mathrm{pre}}, y_{\mathrm{pre}})$ is the predicted track coordinate, and $(x_{\mathrm{Tr}}, y_{\mathrm{Tr}})$ is the true value of the track. The basic prediction process uses 10 historical time steps of real data to predict a future time step. In predicting multiple future time steps, the predicted values are combined with the last 9 of the previous 10-time step data to form a new 10-time step data, which is used as an input for prediction.

The results of calculating the average distance error by predicting different time steps are shown in Table 1. It indicates that Att-BiLSTM is better than the classical short-time memory network under the same prediction time step. However, with the increase of the prediction time step, the prediction accuracy is also declining, but the track direction of the subject is still accurate. As shown in Fig. 8, prediction_1 predicts one-time step in the future.

Table 2 shows the influence of different historical sequence lengths on the prediction results. It indicates that when

Fig. 8. Comparison of different forecasting time series.

Table 2. The influence of different historical sequence lengths on the prediction results.

Evaluation index	Prediction time step	Historical sequence lengths		
		5	10	20
ADE	1	8.261	12.343	6.675
	2	13.328	19.699	31.138
	5	47.278	58.163	109.880

predicting the future trajectory information, it is not that the longer the input historical sequence is, the better. If the input historical sequence is too long, the prediction accuracy may decrease, so the appropriate chronological sequence length should be selected when predicting the future trajectory of surrounding motor vehicles. Although the average distance error of some data is very small, there are great differences in the visual effect of the resulting graph, and it is speculated that there may be over-fitting.

4.5 Experimental comparative analysis

Based on a comparison with LSTM, a further experimental comparative analysis was carried out. The objects of comparison are ConvLSTM and TrajGRU. LSTM is the stacking and time connection of a plurality of LSTM units. ConvLSTM has a convolution structure in input-to-state and state-to-state transition, and a coding prediction structure is formed by

DOI: 10.52396/JUSTC-2021-0155
JUSTC, **2022**, 52(9): 6

stacking multiple ConvLSTM layers. It is roughly composed of two networks: a coding network and a prediction network. The initial state of the prediction network is copied from the last network of the coding network[23]. TrajGRU uses the current and previous input states to generate local neighbor sets of each location at each time tamp and can learn the connection topology by learning the parameters of the subnet. The model inserts a down-sampling layer and an up-sampling layer between neural networks, which are realized by convolution and de-convolution with step size[24].

Both ConvLSTM and TrajGRU are classical algorithms for spatiotemporal sequence prediction. And they have a good effect and strong adaptability. At the same time, they put forward a new comprehensive evaluation scheme to verify whether the new algorithm is SOTA. Compared with them, it can show the advantages of the presented algorithm.

Table 3 describes the performance of different models, including this model on the NGSIM dataset. By analyzing the performance of each model in the next 1 to 5 time steps, we get the mean square error of their predicted coordinates and real coordinates at different time steps. Based on the LSTM model, the improvement of other models relative to LSTM is analyzed. It shows that this model has a better prediction effect than other models under the same conditions.

On the basis of Table 3, take the future forecast time step as the horizontal coordinate axis and the mean square error between the forecast result and the real result as the coordinate vertical axis. The relative results of different models are shown in Fig. 9. It indicates that the prediction effect of LSTM is the worst. Our model works best when the prediction time step is small, while ConvLSTM and TrajGRU are comparable. When the prediction time step is 3, the prediction error of ConvLSTM is close to that of our model. When the prediction time step is 4, the prediction error of TrajGRU is close to that of our model. When the prediction step size is 5, the accuracy of the TrajGRU model is higher than ours, but our model's the total accuracy of the first five prediction steps is still higher than other models. We can say that the model in this work takes into account the advantages of ConvLSTM and TrajGRU in a small prediction time step.

Table 3. The performance of different models.

Models	time step: 1	Δ (ft)	time step: 2	Δ (ft)	time step: 3	Δ (ft)	time step: 4	Δ (ft)	time step: 5	Δ (ft)
LSTM	23.072	—	36.764	—	53.280	—	78.564	—	107.384	—
ConvLSTM	20.150	−2.922	22.495	−14.269	26.620	−26.660	54.459	−24.105	66.248	−41.136
TrajGRU	18.112	−4.960	24.730	−12.034	31.130	−22.150	38.372	−40.192	47.150	−60.234
ours	12.343	−10.729	19.699	−17.065	25.385	−27.895	33.621	−44.943	58.163	−49.221

Fig. 9. The mean square error between the forecast result and the real result of each model.

Fig. 10. The increase of prediction time step on the interference degree of each model.

Fig. 10 gives the influence of the increase of prediction time step on the interference degree of each model more intuitively. The growth of prediction time step has little influence on LSTM and TrajGRU, but it has some influence on ConvLSTM and our model. Specifically, when the prediction time step increases from 3 to 4, the error of ConvLSTM fluctuates greatly, and from 4 to 5, the error of our model also fluctuates slightly. However, our model still has great advantages in overall error.

5 Conclusions and future work

This work investigates the problem of the surrounding vehicles' behavior prediction. By adopting the attention mechanism into the encoder-decoder of a bidirectional long-short memory network and giving different weights to different states according to the importance of the network training process, we solve the problem that the encoder will forget if the sequence is too long. We present detailed comparisons to show the high accuracy of the method.

We successfully solved the contribution of the different information of each time period in the input sequence to the final prediction result. However, some issues, such as, complex environments and multi-vehicle interaction problems, need to be researched in the future.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (U20A20225, U2013601), the Funda-

DOI: 10.52396/JUSTC-2021-0155
JUSTC, **2022**, 52(9): 6

mental Research Funds for the Central Universities, the Natural Science Foundation of Hefei (2021032), the Key Research and Development Plan of Anhui Province (202004a05020058), the CAAI-Huawei MindSpore Open Fund, and University of Science and Technology of China-NIO Intelligent Electric Vehicle Joint Project.

Conflict of interest

The authors declare that they have no conflict of interest.

Biographies

Yunqing Gao received his B.S. degree in Automation from the University of Science and Technology of China (USTC) in 2020. He is currently pursuing the M.S. degree in Control Engineering at USTC. His research interests include automatic control and trajectory prediction.

Hongbo Gao received his Ph.D. degree from Beihang University in 2016. He is currently an associate professor at the Department of Automation in the University of Science and Technology of China. He is the author or coauthor of over 40 journal papers, and the co-holder of 10 patent applications. His research interests include unmanned system platform, robotics, machine learning, decision support system, and intelligent driving.

References

[1] Li D, Gao H. A hardware platform framework for an intelligent vehicle based on a driving brain. *Engineering,* **2018**, *4* (4): 464–470.

[2] Lefèvre S, Vasquez D, Laugier C. A survey on motion prediction and risk assessment for intelligent vehicles. *Robomech Journal,* **2014**, *1* (1): 1–14.

[3] Brännström M, Coelingh E, Sjöberg J. Model-based threat assessment for avoiding arbitrary vehicle collisions. *IEEE Transactions on Intelligent Transportation Systems,* **2010**, *11* (3): 658–669.

[4] Ortiz M G, Fritsch J, Kummert F, et al. Behavior prediction at multiple time-scales in inner-city scenarios. In: 2011 IEEE Intelligent Vehicles Symposium (IV). Baden-Baden, Germany: IEEE, **2011**: 1068−1073.

[5] Xie G, Gao H, Qian L, et al. Vehicle trajectory prediction by integrating physics-and maneuver-based approaches using interactive multiple models. *IEEE Transactions on Industrial Electronics,* **2018**, *65* (7): 5999–6008.

[6] Zheng X, Zhang D, Gao H, et al. A novel framework for road traffic risk assessment with HMM-based prediction model. *Sensors,* **2018**, *18* (12): 4313.

[7] Gao H, Cheng B, Wnag K, et al. Object classification using CNN-based fusion of vision and LIDAR in autonomous vehicle environment. *IEEE Transactions on Industrial Informatics,* **2018**, *14* (9): 4224–4231.

[8] Gao H, Su H, Cai Y, et al. Trajectory prediction of cyclist based on dynamic Bayesian network and long short-term memory model at unsignalized intersections. *Science China Information Sciences,* **2021**, *64*: 172207.

[9] Pu J, Zhou W, Hu H, et al. Boosting continuous sign language recognition via cross modality augmentation. In: Proceedings of the 28th ACM International Conference on Multimedia. Seattle, USA: Association for Computing Machinery, **2020**: 1497–1505.

[10] Ondruska P, Posner I. Deep tracking: Seeing beyond seeing using recurrent neural networks. *Proceedings of the AAAI Conference on Artificial Intelligence,* **2016**, *30* (1): 3361–3367.

[11] Khosroshahi A, Ohn-Bar E, Trivedi M M. Surround vehicles trajectory analysis with recurrent neural networks. In: 2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC). Piscataway, USA: IEEE, **2016**: 2267–2272.

[12] Jeong Y, Kim S, Yi K. Surround vehicle motion prediction using LSTM-RNN for motion planning of autonomous vehicles at multi-lane turn intersections. *IEEE Open Journal of Intelligent Transportation Systems,* **2020**, *1*: 2–14.

[13] Park S H, Kim B D, Kang C M, et al. Sequence-to-sequence prediction of vehicle trajectory via LSTM encoder-decoder architecture. In: 2018 IEEE Intelligent Vehicles Symposium (IV). Jiangsu, China: IEEE, **2018**: 1672–1678.

[14] Benterki A, Judalet V, Maaoui C, et al. Multi-model and learning-based framework for real-time trajectory prediction. In: 2020 28th Mediterranean Conference on Control and Automation. Akko, Israel: IEEE, **2020**: 776–781.

[15] Gao H, Guo F, Zhu J, et al. Human motion segmentation based on structure constraint matrix factorization. *Science China Information Sciences,* **2022**, *65*: 119103.

[16] Dai S, Li L, Li Z. Modeling vehicle interactions via modified LSTM models for trajectory prediction. *IEEE Access,* **2019**, *7*: 38287–38296.

[17] Xin L, Wang P, Chan C Y, et al. Intention-aware long horizon trajectory prediction of surrounding vehicles using dual LSTM networks. In: 2018 21st International Conference on Intelligent Transportation Systems (ITSC). Hawaii, USA: IEEE, **2018**: 1441–1446.

[18] Guan D, Zhao H, Zhao L, et al. Intelligent prediction of mobile vehicle trajectory based on space-time information. In: 2019 IEEE 89th Vehicular Technology Conference (VTC2019-Spring). Kuala Lumpur, Malaysia: IEEE, **2019**: 1–5.

[19] Gao H, Zhu J, Zhang T, et al. Situational assessment for intelligent vehicles based on stochastic model and Gaussian distributions in typical traffic scenarios. *IEEE Transactions on Systems, Man, and Cybernetics: Systems,* **2022**, *52* (3): 1426–1436.

[20] Jeong Y, Yi K. Bidirectional long shot-term memory-based interactive motion prediction of cut-in vehicles in urban environments. *IEEE Access,* **2020**, *8*: 106183–106197.

[21] Lenz D, Diehl F, Le M T, et al. Deep neural networks for Markovian interactive scene prediction in highway scenarios. In: 2017 IEEE Intelligent Vehicles Symposium (IV). Los Angeles, USA: IEEE, **2017**: 685–692.

[22] Deo N, Trivedi M M. Multi-modal trajectory prediction of surrounding vehicles with maneuver based LSTMs. In: 2018 IEEE Intelligent Vehicles Symposium (IV). Changshu, China: IEEE, **2018**: 1179–1184.

[23] Shi X, Chen Z, Wang H, et al. Convolutional LSTM network: A machine learning approach for precipitation nowcasting. In: Proceedings of the 28th International Conference on Neural Information Processing Systems-Volume 1. Montreal, Canada: MIT Press, **2015**: 802–810.

[24] Shi X, Gao Z, Lausen L, et al. Deep learning for precipitation nowcasting: A benchmark and a new model. [2022-06-15]. https://arxiv.org/abs/1706.03458.

DOI: 10.52396/JUSTC-2021-0155
JUSTC, **2022**, 52(9): 6