Prediction Method of Multi-Injection Pressure Fluctuation of Diesel Engine Based on Recurrent Neural Network Model

Zhe Zuo[*1], Yu Zhang[*2], and Meng Du[*3]

[*1*2*3]School of Mechanical Engineering, Beijing Institute of Technology, Beijing 100081, China;
[*1]E-mail: zuzeus@bit.edu.cn
[*2]E-mail: zhangyu1088@163.com

Abstract. In the high-pressure common rail fuel injection system, the pressure fluctuation caused by single injection in one cycle makes the injection pressure unstable, which reduces the control accuracy and makes it difficult to determine the circulating fuel supply. In recent years, with the development of deep learning, it provides a new idea for the prediction of pressure fluctuation. The general time series model and neural network model are difficult to meet the accuracy requirements. This paper proposes a algorithm: using LSTM and Seq2Seq as the infrastructure, introducing the Global Attention Mechanism, and optimizing by using Probabilistic Teacher Forcing, Adam and other strategies. The performance of the new model in different target injection pressure, different injection duration and different pre-main injection time interval is analyzed.

Keywords: Neural Network,Pressure Fluctuation, LSTM,Seq2Seq,Attention Mechanism

1. INTRODUCTION

With the development of multiple injection technology, researchers found that the fuel quantity of the second or third injection is different from the expected fuel quantity[1]. Different from single injection, the interval between two adjacent injections is small, and the previous injection will affect the next injection. As diesel is a viscous fluid, the subsequent fuel can not be replenished immediately after injection, so the pressure fluctuation in the injector is inevitable[2-3]. The research on the change law of pressure and time after injection can provide reference for selecting more suitable injection law, optimizing injection efficiency and reducing emission of harmful substances. At present, there are two methods to study pressure fluctuation: bench test and simulation[4].

The nonlinear degree of pressure fluctuation law is very high. On the one hand, it is difficult to predict the law of pressure fluctuation by simple linear regression. On the other hand, deep learning has the advantages of strong nonlinear fitting ability and high generalization.

The essence of deep learning is to learn more useful features by constructing machine learning models with many hidden layers and massive training data, so as to ultimately improve the accuracy of classification or prediction[5]. In the field of engineering, it is necessary to predict the distribution of time series data for a long time under some specific conditions, and the time span of time series data input by the model is large. For this type of time series data, the effect of general ARIMA model or recurrent neural network model is difficult to meet the demand of prediction accuracy[6]. It is of great significance to introduce attention mechanism into traditional engineering field for pressure fluctuation analysis.

At present, there are few researches on the prediction of pressure fluctuation in high-pressure common rail system by machine learning and deep learning, and it is more meaningful to use deep learning model for rail pressure model. In order to solve the shortcomings of traditional numerical simulation methods, this study uses the deep learning method based on neural network to build the model.

2. RECURRENT NEURAL NETWORK

2.1. Structure of LSTM and Seq2Seq Unit

It is unreasonable to build the prediction model of ordinary neural network, because the prediction of pressure fluctuation is essentially a time series problem, and the ordinary neural network structure can not learn the time information hidden in the time series data. The general RNN structure, such as Elman network, is relatively simple. In structure, due to the use of one-step descent method for model training, Elman network will have serious gradient explosion or gradient disappearance problems when the sequence length is large and the network structure is complex[7-8]. Therefore, the prediction model is built based on the structure of Seq2Seq with LSTM as the basic unit.

The network unit structure of LSTM is shown in Fig. 1.

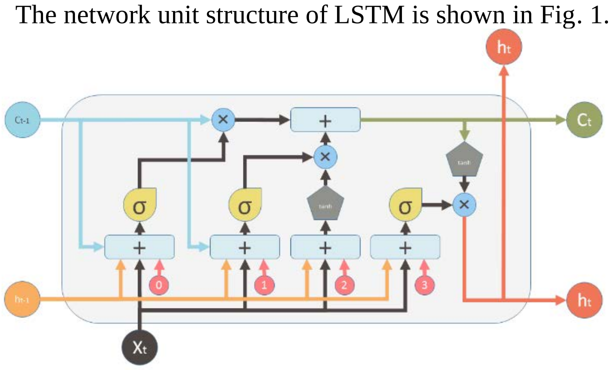


Fig. 1 LSTM network unit structure

The main improvement of LSTM is that three gating units are added to the original RNN network: input gate i_t, Output gate o_t, Forgetting gate f_t. In the forward process, not only the hidden state h_t is passed between time steps. In addition, a "peephole" connection channel is added to transfer the cell memory value c_t, W_{ix}、W_{fx}、W_{ox}、W_{gx}、b_{ix} and b_{ih}.The equivalent represents the weight and bias of each gating unit[9-11].

The structure of Seq2Seq was proposed by Google, and it was first used in the fields of machine translation and voice conversion[12]. The structure of Seq2Seq consists of two parts: encoder and decoder. The structure is shown in Fig. 2. The encoder is responsible for compressing the input sequence into a text vector with a fixed dimension size, and then passing the text vector to the decoder. The decoder decodes the information contained in the text vector and outputs the prediction sequence step by step according to the time sequence. The encoder and decoder are composed of basic RNN units.

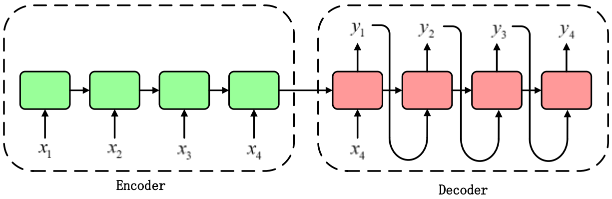


Fig. 2 Structure of Seq2Seq

2.2. The Principle of Attention Mechanism

When the encoder compresses the information in the pre-injection, the text vector is difficult to include all the pre spray information due to the constant dimension of the text vector, and the pressure fluctuation information in the early stage before the pre-injection is seriously lost; when the decoder decodes the text vector, due to the information loss of the text vector, the model cannot obtain enough effective information in the pre spray, and the final model prediction accuracy is poor[13]. Considering the formation mechanism of pressure fluctuation, the prediction model should have obtained enough effective information of pressure fluctuation before prediction. Therefore, attention mechanism is introduced into the original structure of Seq2Seq based on LSTM.

At present, attention mechanism has been derived from a variety of structures and applied to various algorithms in various fields. The structure of attention mechanism is shown in Fig. 3.

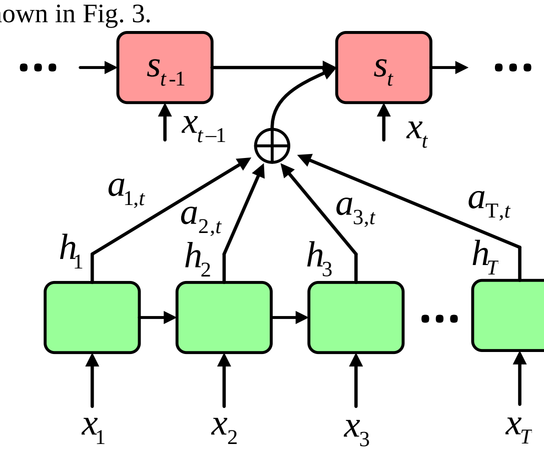


Fig. 3 The structure of attention

Firstly, in the process of decoding the text vector, the similarity between the current hidden state of the decoder s_{t-1} and the hidden state $(h_1,\ h_2, \dots ,\ h_T)$ in each time step of the encoder is calculated to obtain the corresponding score of the current decoder hidden state to each hidden state in the encoder. Then, the normalized exponential function (Softmax) is used to regularize the score. Softmax can normalize the elements to (0,1). The normalized fraction is multiplied by the hidden state of the encoder to obtain the content vector c_t. Finally, the content vector c_t is input into the RNN unit of the current decoder for final output.

Through the introduction of global attention mechanism, the information obtained in decoder is not only limited to fixed dimension text vector, but also can get the hidden information of each position of input sequence through content vector, so as to pay attention to the position more related to its own output and improve the prediction effect of the model.

During the duration of main injection, the law of pressure fluctuation is closely related to the rated injection pressure at the beginning of pre-injection and the drop degree of needle valve opening pressure. Through the attention mechanism, the model can automatically learn the related information, thus assisting the model learning and improving the model effect.

2.3. Model Structure Design

Before the current information is input into the model, an embedded layer is added to smooth the input information and improve the nonlinearity of the model. In the decoder stage, the global attention mechanism is combined to predict the decoding. The model built by the above process is referred to as ATTN-LSTM model, and the model structure is shown in Fig. 4.

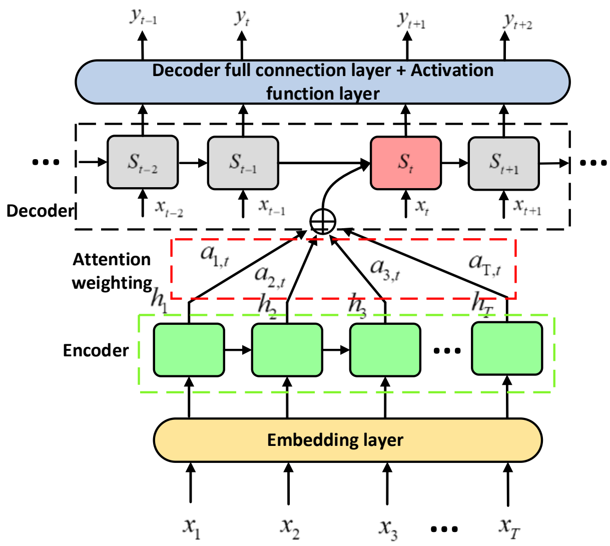


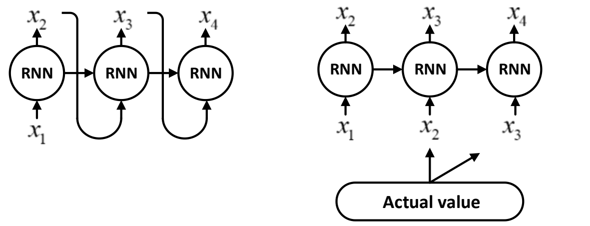
Fig. 4 Network structure of prediction model

3. BASIC PPTIMIZATION STRATEGY

3.1. Probabilistic Teacher Forcing Training Strategy

Due to the high complexity of ATTN-LSTM model, in order to speed up the model convergence speed and save the calculation cost; on the other hand, in order to avoid the model falling into local optimum, optimize the model optimization path and improve the final prediction accuracy of the model, this section will optimize the training strategy and optimizer of the model.

Teacher forcing strategy is a method for training time series model, which can effectively speed up the convergence speed of the model and improve the accuracy of the model. The principle of teacher forcing is to change the input received by each RNN unit from the output value of the previous time step to the target value of the previous time step, as shown in Fig. 5.



(a) No teacher forcing strategy (b) Teacher forcing strategy

Fig. 5 Teacher forcing schematic diagram

Probabilistic teacher forcing strategy is adopted. In the training process of the model, the teacher forcing is set to 100%, and the model receives the real value of the previous time step in each iteration process; as the error gradually decreases, the teacher forcing probability is reduced dynamically; when the error is reduced to a certain level, the teacher forcing strategy is completely closed to enable the model to adjust the optimization direction independently.

3.2. Early Stop Technology

In order to avoid the model from entering the over fitting area, improve the generalization ability of the model, and effectively save the training time, this study adopts the early stop strategy. Among them, generalization ability refers to the prediction ability of neural network for untrained samples.

In the process of model training, the error between the training set and the test set will gradually decrease. When the model converges to a certain degree, while the error of the training set continues to decrease, the error of the test set begins to increase, as shown in Fig. 6, where the horizontal axis is the number of iterations and the vertical axis is the current error. At this time, the model has over fitting phenomenon, so the model should terminate the training before the test set error begins to rise, and save the current optimal model.

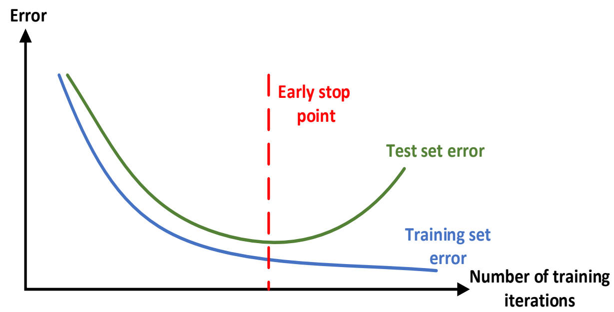


Fig. 6 Schematic diagram of early stop strategy

3.3. Adam Optimizer

Adam algorithm is an adaptive learning rate algorithm, which uses the momentum mechanism for reference, adaptively changes the learning rate, speeds up the convergence speed, and reduces the oscillation when the gradient drops.

4. EXPERIMENT AND VERIFICATION

4.1. Parameter Setting and Experimental Evaluation

The parameters used in the experiment are shown in Table. 1.

Table. 1 Experimental parameter table

Parameter name	Parameter value
Number of training set samples	48000
Number of test set samples	12000
Enter sequence length	20
Predicted sequence length	100
Learning rate	0.005
Batch size	6
Feature dimension	13
Optimizer	Adam
Loss function	MSE
Evaluation index	MSE、MAPE

The data flow of pre-injection pressure fluctuation information into the network until the final result is output is shown in Table. 2.

The table shows the process that the encoder encodes the pre spray fluctuation information into a text vector and decodes the text vector at the decoder and predicts the pressure fluctuation information of the first time step. Because of the particularity of RNN structure, it is necessary to input information step by step when predicting the information. The whole pressure fluctuation process can be predicted by repeating the decoder data stream for 100 times. The flow chart of model training is shown in Fig. 7.

The bench test of high pressure common rail system in this research is mainly to collect the original data by changing the rated injection pressure, spraying time interval and spraying duration under the condition of unchanged system structure. In order to obtain the pressure fluctuation during the pre-injection period by the parameters such as the duration of pre-injection and the rated injection pressure, the generation model of pre-injection pressure fluctuation was established by using the common shallow neural network. Because the internal pressure fluctuation law of pre-injection collected in this experiment is relatively simple, it is not suitable to use too complex model to generate, so the generation model of this study adopts the above-mentioned simple shallow neural network.

In order to verify the effectiveness of the prediction model established in this chapter, the pressure fluctuation problem in multiple injections of high pressure common rail system was compared with different model structures and training strategies, and the results were analyzed. The experimental results of each model are shown in the Table. 3.

Table. 2 Data flow

Layer definition	Parameter	Data flow dimension
Embedding layer	Number of neuron nodes: 56	Input:（6，20，13）
		Output:（6，20，13）
Encoder LSTM layer	Number of hidden state nodes: 512 LSTM layers: 1	Input: (1). Sequence embedded information: (6, 20, 13) (2). Hidden state set: (6, 20, 512)
		Output:（6， 20， 512）
Decoder LSTM layer	Number of hidden state nodes: 512 LSTM layers: 1	Input: Text vector: (6, 1, 512)
		Output: (6, 1, 512)
Decoder full connection layer 1	Number of neuron nodes: 56	Input:（6，1，512）
		Output:（6，1，56）
Nonlinear activation layer Activation function (Relu)	NONE	Input:（6，1，56）
		Output:（6，1，56）
Decoder full connection layer 2	Number of neuron nodes: 1	Input:（6，1，56）
		Output:（6，1，1）
Nonlinear activation layer Activation function (Sigmoid)	NONE	Input:（6，1，1）
		Output:（6，1，1）

In the table, the first model is the Seq2Seq model with Elman network as the unit structure, and the basic gradient descent algorithm is used to optimize the model parameters. Because of its simple structure and fast training speed, but due to the structural defects of the model itself, it rapidly appears over fitting phenomenon in the training set, reaching the termination condition of early stop strategy, stopping training, and the model accuracy is poor;

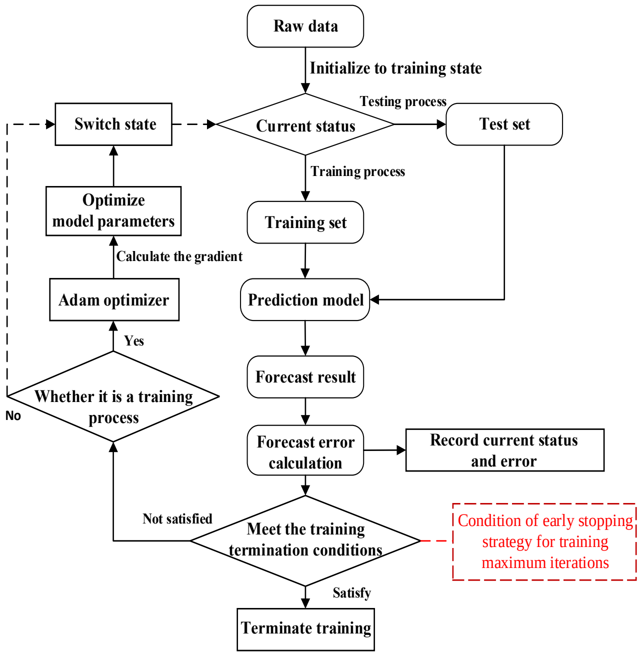


Fig. 7 Training process of prediction model

The second model has the advantages of simple structure and fast training speed In the model, Adam optimizer is added on the basis of the first model. From the training results, the Adam optimizer does not improve the accuracy of the model, but greatly slows down the speed of model over fitting;

The third model is based on LSTM is a unit structured Seq2Seq model, which uses Adam optimizer. Compared with the former two models, the model accuracy is greatly improved, which verifies the effectiveness of LSTM. However, LSTM model is more complex than Elman network, and its training time is longer.

On the basis of the previous model, the fourth model adds the global attention mechanism, which greatly improves the accuracy of the model, reflecting the important role of attention mechanism. The prediction accuracy has basically met the demand of pressure fluctuation prediction. Due to the addition of attention mechanism, the complexity of the model is improved and the single step training time is longer.

Table. 3 Model effect

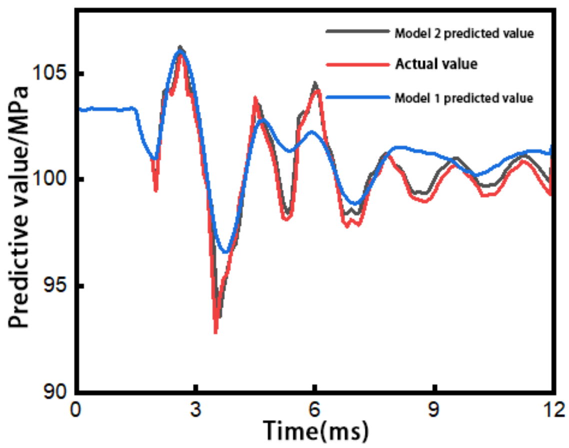
Model name	Step training time (s)	Overall training time (min)	Iteration steps	Training Set MSE	Test Set MSE	Test Set MAPE
Seq2Seq(Elman)	0.041	5.04	7317	9.8e-3	1.1e-2	2.984
Seq2Seq(Elman)+Adam	0.044	8.72	11890	1.1e-2	1.2e-2	3.174
Seq2Seq(LSTM)+Adam	0.088	31.53	21497	5.4e-4	7.1e-4	1.465
ATTN+Seq2Seq(LSTM)+ Adam	0.132	41.23	18742	2.4e-5	4.2e-5	0.653
ATTN+Seq2Seq(LSTM)+ Adam+ Probability TF	0.143	31.33	13145	3.2e-5	4.7e-5	0.416

The fifth model is the ATTN-LSTM model used in this study. On the basis of the fourth model, the probabilistic teacher forcing strategy is added to accelerate the overall convergence speed of the model, and reduce the iteration steps required to achieve the same level of prediction accuracy as the fourth model, which verifies the effectiveness of the teacher forcing strategy.

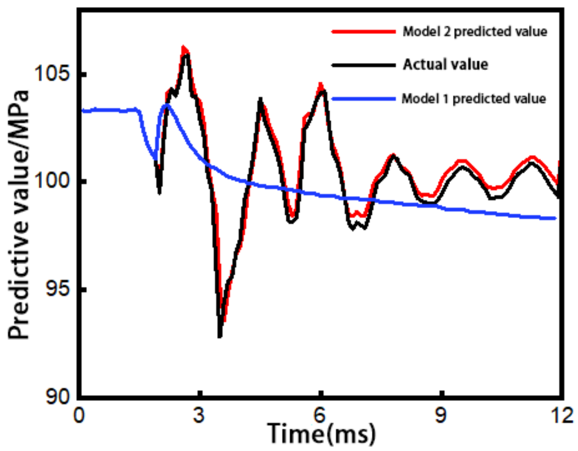
4.2. Experimental Results and Analysis

In order to make the prediction effect of the model more intuitive, we compared the results of the second model, the third model and the fifth model visually, as shown in Fig. 8. The Elman network used in model 2 is difficult to meet the demand of rail pressure prediction, and the model is seriously under fitted. Model 3 can basically predict the change trend of pressure fluctuation, but the accuracy is poor.

Through comparison, it is found that both LSTM structure and global attention mechanism can effectively improve the prediction accuracy of the model.



(a)Injection interval 0.5ms

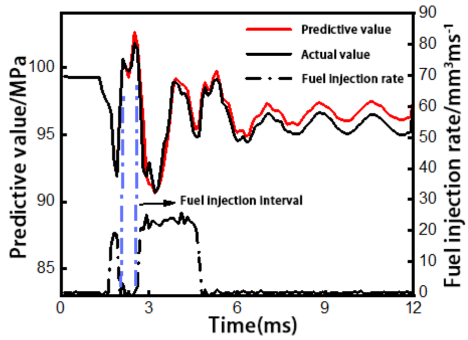


(b)Injection interval 1.0ms

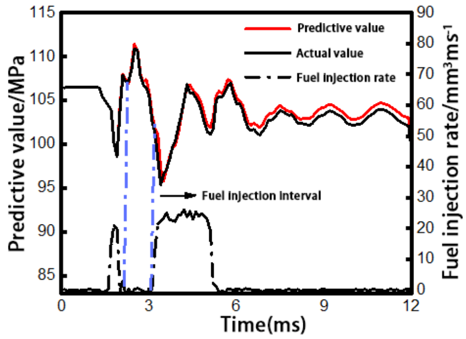
Fig. 8 Model effect comparison

In order to further verify the generalization of ATTN-LSTM built in this chapter, the results of different injection time interval(time interval between pre-injection and main-injection) and different rated injection pressure were visualized.

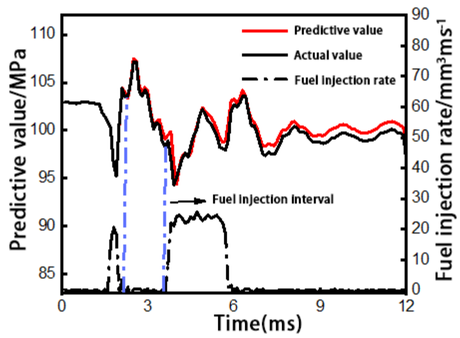
Fig. 9 shows the prediction results under different injection time intervals under the rated injection pressure of 100MPa.



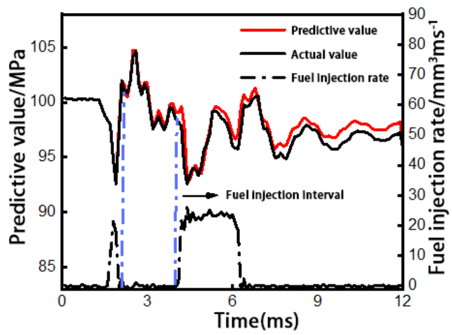
(a) Interval 0.5ms



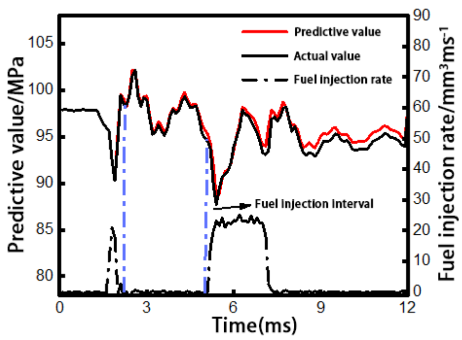
(b) Interval 1.0ms



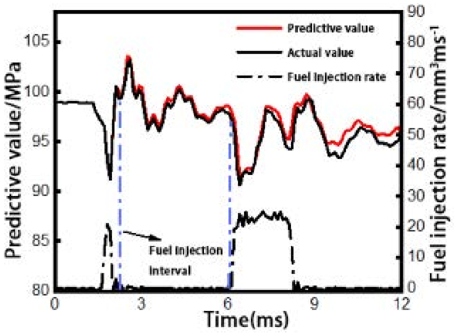
(c) Interval 1.5ms



(d) Interval 2.0ms



(e) Interval 3.0ms

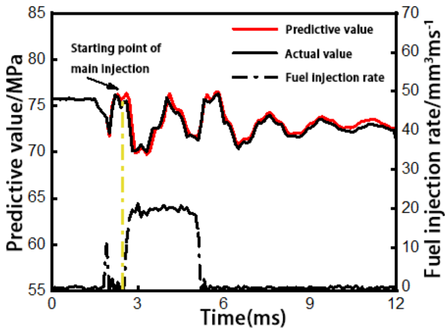


(f) Interval 4.0ms

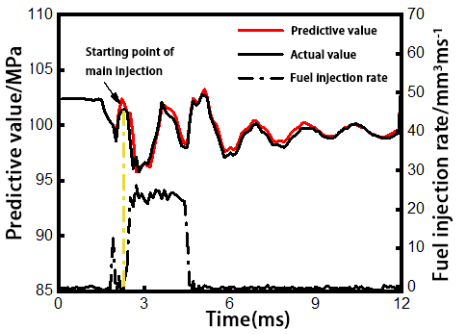
Fig. 9 Shows the prediction results under different injection time intervals between pre-injection and main-injection under the rated injection pressure of 100MPa

The average prediction accuracy of pressure fluctuation in the figure is 99.5%, which indicates that ATTN-LSTM can accurately predict pressure fluctuation under different injection time intervals, and has strong generalization for different injection time intervals.

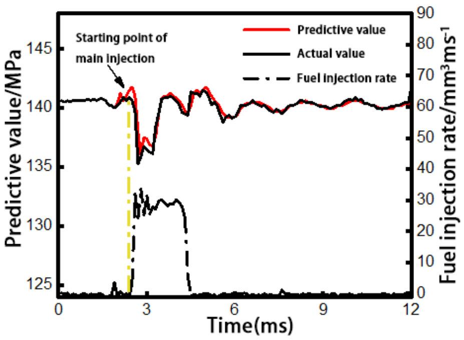
The prediction results of the model with injection time interval of 0.5ms and different rated injection pressure are visualized, as shown in Fig. 10.



(a) 80MPa



(b) 100MPa



(c) 140MPa

Fig. 10 Prediction results of the model with injection interval of 0.1ms and different target injection pressure

The prediction accuracy of the model is high under different rated injection pressure, and the predicted value is in good agreement with the real pressure fluctuation curve, which proves that the generalization performance of the model at different rated injection pressure is strong.

To sum up, it is found that the ATTN-LSTM model established in this study can meet the demand of pressure fluctuation prediction with high accuracy and good generalization performance. It is suitable for various working conditions such as different target injection pressure, different injection duration and different injection time interval.

5. CONCLUSION

This paper optimizes the model structure selection, network structure design and model training strategy to meet the demand of pressure fluctuation prediction. In the selection of model structure, LSTM and Seq2Seq are selected as the basic structures. In order to improve the prediction accuracy of the model, the global attention mechanism is added to the original structure, and the ATTN-LSTM model is designed. In the network structure design, considering the model accuracy and training calculation burden and other factors, a more reasonable network structure is designed. In terms of training strategy, the training strategies such as probability teacher forcing and early stop are adopted. The Adam optimizer adaptively changes the learning rate, speeds up the convergence speed and reduces the oscillation when the gradient drops.

The focus of this study is to use the pre-injection information to predict the more complex and longer duration of subsequent pressure fluctuations. The pre-injection pressure fluctuation information generated in this paper will be used as the front-end auxiliary model of the subsequent pressure fluctuation prediction model, and the pre-injection pressure fluctuation sequence will be generated as the input and transmitted to the subsequent model.

ACKNOWLEDGEMENT

The authors gratefully acknowledge the financial support by Beijing Natural Science Foundation(grant no. 3182033)

REFERENCES

[1] N. A. Henein, M-C. Lai, I. P. Singh,etc.Characteristics of a Common Rail Diesel Injection System under Pilot and Post Injection Modes[R].Detroit, Michigan, USA:SAE 2002 World Congress,2002.

[2] Catania A E , Ferrari A , Manno M , et al. Experimental Investigation of Dynamics Effects on Multiple-Injection Common Rail System Performance[J]. Journal of Engineering for Gas Turbines & Power, 2008, 130(3):460-466.

[3] Stefano Ubertini.Injection pressure fluctuations model applied to a multi-dimensional code for diesel engines simulation[C].Biennial Conference on Engineering Systems Design & Analysis.Manchester, United Kingdom:American Society of Mechanical Engineers,2004.437-445.

[4] Zhu Keqing, Xu Quankui, Yang Lin, et al. Multiple-injection quantity measurement and analysis in high pressure common rail injection system[J]. Transactions of the CSAE, 2007, 23(8): 117-121.

[5] Yu Kai, Jia Lei, Chen Yuqiang, Xu Wei. Yesterday, today and tomorrow of deep learning [J]. Computer research and development, 2013,50 (09): 1799-1804

[6] Menggang Li,Changsheng Zhou, Lian Lian,etc.Simulation research of industrial enterprise total profits based on the neural network of ARIMA[R].:2016 International Conference on Industrial Economics System and Industrial Security Engineering,2016.

[7] Hochreiter S , Schmidhuber J . Long Short-Term Memory[J]. Neural Computation, 1997, 9(8):1735-1780.

[8] Wang Xin, Wu Ji, Liu Chao, et al. Exploring LSTM based recurrent neural network for failure time series prediction[J]. Journal of Beijing University of Aeronautics and Astronautics, 2018, 44 (04): 772-784

[9] Elman J L. Finding structure in time [J]. Cognitive Science, 1990, 14(2): 179-211.

[10] G E, Hinton,R R, Salakhutdinov.Reducing the dimensionality of data with neural networks.[J].Science (New York, N.Y.),2006,313(5786):504-7.

[11] Sammut C , Webb G I.Machine Learning[M].Springer,2010:139-13 9.

[12] Vincent Le Guen,Nicolas Thome.Shape and Time Distortion Loss for Training Deep Time Series Forecasting Models[R].Vancouver, Canada.:33rd Conference on Neural Information Processing Systems,2019.

[13] Sutskever I , Vinyals O , Le Q V . Sequence to Sequence Learning with Neural Networks[J]. Advances in neural information processing systems,arXiv:1409.3215, 2014