Mechanical & Materials Engineering Faculty Publications

Mechanical & Materials Engineering, Department of

6-12-2020

LSTM-Based Anomaly Detection for Non-Linear Dynamical System

Yue Tan

Chungjing Hu

Kuan Zhang

Kan Zeng

Ethan A. Davis

See next page for additional authors

Authors

Yue Tan, Chungjing Hu, Kuan Zhang, Kan Zeng, Ethan A. Davis, and Jae Sung Park

LSTM-Based Anomaly Detection for Non-Linear Dynamical System

YUE TAN [1], (Graduate Student Member, IEEE), **CHUNJING HU** [2],
KUAN ZHANG [3], (Member, IEEE), **KAN ZHENG** [1], (Senior Member, IEEE),
ETHAN A. DAVIS [4], **AND JAE SUNG PARK** [4]

[1] Intelligent Computing and Communication (ICC) Laboratory, Beijing University of Posts and Telecommunications, Beijing 100876, China
[2] Key Laboratory of Universal Wireless Communications, Beijing University of Posts and Telecommunications, Beijing 100876, China
[3] Department of Electrical and Computer Engineering, University of Nebraska-Lincoln, Lincoln, NE 68588, USA
[4] Department of Mechanical and Materials Engineering, University of Nebraska-Lincoln, Lincoln, NE 68588, USA

Corresponding author: Chunjing Hu (hucj@bupt.edu.cn)

ABSTRACT Anomaly detection for non-linear dynamical system plays an important role in ensuring the system stability. However, it is usually complex and has to be solved by large-scale simulation which requires extensive computing resources. In this paper, we propose a novel anomaly detection scheme in non-linear dynamical system based on Long Short-Term Memory (LSTM) to capture complex temporal changes of the time sequence and make multi-step predictions. Specifically, we first present the framework of LSTM-based anomaly detection in non-linear dynamical system, including data preprocessing, multi-step prediction and anomaly detection. According to the prediction requirement, two types of training modes are explored in multi-step prediction, where samples in a wall shear stress dataset are collected by an adaptive sliding window. On the basis of the multi-step prediction result, a Local Average with Adaptive Parameters (LAAP) algorithm is proposed to extract local numerical features of the time sequence and estimate the upcoming anomaly. The experimental results show that our proposed multi-step prediction method can achieve a higher prediction accuracy than traditional method in wall shear stress dataset, and the LAAP algorithm performs better than the absolute value-based method in anomaly detection task.

INDEX TERMS LSTM, anomaly detection, non-linear dynamical system, multi-step prediction, time series.

I. INTRODUCTION

Dynamical system is the basic framework for modeling and control of an enormous variety of complicated systems, including fluid dynamics, signal propagation and interference in electronic circuits, heat transfer, biological systems, chemically reacting flows, etc. [1]. Nonlinearity is an important feature of these complex dynamical systems, as a result of a rich diversity of observed dynamical behaviors across the physical, biological, and engineering sciences. Modern non-linear dynamical systems are becoming more and more complex, where a variety of complicated phenomena make them vulnerable to software and hardware problems, causing anomalies in various emerging applications. It is necessary to predict and recover anomalies in time. However, in practice, this task is still manually solved. It is essential to provide

automatic anomaly prediction methods for non-linear dynamical systems.

As time goes by, a large amount of time series data are produced by non-linear dynamical system. Some of them can be obtained by non-linear mapping and derivation of differential equations [2]. During the past few decades, some techniques, such as Large-Eddy Simulation (LES) and Reynolds-averaged Navier–Stokes (RANS), were proposed to predict turbulent flow in grid-resolved scales accurately. However, LES requires a dedicated model for the effect on grid-resolved quantities [3], [4]. RANS also need to model the turbulence first in a temporally averaged sense [5], [6]. To raise the advance anomaly alerts, non-linear dynamical systems should be continuously monitored.

Although the non-linear governing equations are usually known, simulations have to take extended periods of time and become computationally expensive, and time resolution with a certain accuracy is needed [7]. Deep learning is a useful method for modeling analysis and achieves a good

The associate editor coordinating the review of this manuscript and approving it for publication was Hamid Mohammad-Sedighi.

result in many tasks including video classification, speech recognition, and natural language processing [8], [9]. Deep learning methods also obtain results with high accuracy in complex prediction problems. However, there are few results based on deep learning methods in the field of non-linear dynamical system modeling.

We propose an LSTM-based method to the anomaly prediction task for non-linear dynamical system. Compared with traditional prediction methods, the LSTM-based method achieves a higher prediction accuracy for different zones of the time series. In particular, by incorporating LSTM into the developed anomaly detection algorithm, the temporal features of time series data are extracted in order to predict multi-step wall shear stress value of non-linear system. The main contributions of this paper lie in the following aspects.

- We identify the problem of predicting multi-step wall shear stress and conduct experiments in a non-linear dynamical system. An LSTM-based anomaly detection method is proposed to solve this prediction problem. The proposed scheme employs a multi-layer network based on LSTM units, which captures complex temporal influences and pickup-drop-off interactions effectively.
- The anomaly points that reflect the latent danger to the system are identified. An effective anomaly detection algorithm, named Local Average with Adaptive Parameters (LAAP), is developed to exploit the anomalous period on predicted data. This algorithm is easily incorporated into the prediction model, to boost the performance of anomaly detection.
- Extensive experiments are conducted to evaluate the performance of the proposed methods. The dataset contains time series value of wall shear stress collected from a fluid non-linear dynamical system. The results show that our method achieves a higher prediction accuracy in both multi-step prediction and anomaly detection than a typical inference algorithm called Autoregressive Integrated Moving Average (ARIMA).

The remainder of this paper is organized as follows. Section II introduces the related work on LSTM-based time series prediction and anomaly detection. Section III presents a detailed description of the proposed prediction methods and anomaly detection algorithm. In Section IV, the experimental results and analysis depending on the proposed approach are presented. Finally, conclusions are given in Section V.

II. RELATED WORK

Time series prediction and anomaly detection have been widely studied and applied to a variety of real-world projects. In this section, related works on anomaly detection are reviewed. Related approaches on neural network for prediction are discussed.

A. ANOMALY DETECTION METHODS

Anomaly detection is a proactive method that raises alerts when the system is still in the normal state but progressing to an anomaly state [10]. To prevent the anomaly or reduce the damage, anomaly detection is of great significance in many fields, such as geology, meteorology, and medicine.

For anomaly detection, the Markov model-based approaches are widely used. A Markov model is a stochastic model used to model randomly changing systems. It can be used to predict the state in the future using previous information. It is able to establish the transition probabilities relationship between states. In Markov chain model, the pattern of different metric values can be recognized and the next state of these values is predicted by the model. For example, Gu *et al.* presented a stream-based mining algorithm by combining Markov models and Bayesian classification methods for online anomaly prediction. Their anomaly prediction scheme can raise alerts for impending system anomalies in advance and suggests possible anomaly causes [11]. Sendi *et al.* proposed a framework to predict multi-step attacks before they pose a serious security risk. They used Hidden Markov Model (HMM) to extract the interactions between attackers and networks [12]. Paulo *et al.* applied a Markov chains-based method to characterize the stochasticity of droughts and predict the transition probability from one class of severity to another up to 3 months ahead [13]. In [14], Zhou *et al.* employed the Evidential Markov chain for the anomaly prediction of PlanetLab. A Belief Markov chain is proposed to extend the Evidential Markov chain and cope with noisy data stream.

Also, there are many anomaly detection approaches based on regression methods. In these approaches, the detection problems are converted into normal regression problems, then machine learning-based regression algorithms and models can be applied to anomaly detection. For instance, Hong et.al proposed a new anomaly detection approach based on principal component analysis (PCA), information entropy theory and support vector regression (SVR). It can be used in credit card fraud detection as well as intrusion detection in cybersecurity [15]. In [16], Huang *et al.* presented a Recurrent Neural Networks (RNN) model for the anomaly prediction of component-based enterprise systems. Their RNN-based method has shown high prediction accuracy and time efficiency for large-scale systems.

B. LSTM FOR PREDICTION

In recent years, neural network-based deep learning approaches are widely applied to the prediction problems [10]. Among them, Recurrent Neural Network (RNN) and its advanced variants have shown a higher performance for prediction tasks than traditional methods. RNN is a type of artificial neural networks. It can capture the feature of input time series data by remembering its historical information. LSTM Network is a special RNN. LSTM is proved to outperform many other types of RNN in modeling sequential data and widely used in prediction tasks [17].

RNN is suitable for learning patterns, relationships and interconnections hidden in time series as well as modeling the temporal sequences. In [18], Zio *et al.* employed Infinite
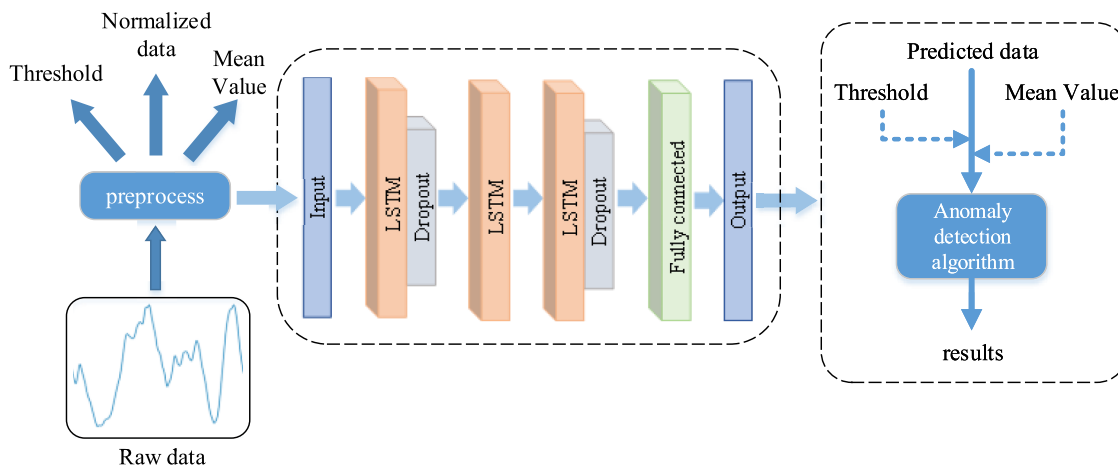
FIGURE 1. The framework of LSTM-based anomaly detection approach.

Impulse Response Locally Recurrent Neural Networks (IIR-LRNN) to forecast failures and make reliability prediction for systems' components. It is innovative to use such dynamic modeling technique in reliability prediction tasks.

Especially, LSTM, which was proposed by Hochreiter and Schmidhuber [19], has emerged to be an effective and scalable type of RNNs for several learning problems related to sequential data [20]. By utilizing multiplicative gates that enforce constant error flow through the internal states of cells, LSTMs overcome the vanishing gradient problem in original RNNs [21].

Due to their superior ability for processing time-series data, LSTMs are widely applied to prediction tasks. In [22], an LSTM-based deep learning method is explored for travel time prediction on the dataset provided by Highways England. They have achieved high prediction accuracy for 1-step ahead travel time prediction error. In [23], an LSTM-based spatio-temporal learning framework is proposed for land cover prediction. The authors design a dual-memory structure to capture both long-term and short-term patterns in temporal sequences. Based on LSTM, the authors in [24] proposed an improved model to learn tweet representation from weakly-labeled data and make tweet classification with higher accuracy. In [17], LSTM network is first used to predict the performance of web servers as the URL requests are always sequential data. The logs of Nginx web servers are analyzed before the performance of web server is predicted.

Besides, Chen *et al.* utilized LSTM for the prediction task for China stock [25]. Altche *et al.* addressed a highway trajectory prediction approach based on LSTM network [26]. Qu *et al.* applied an approach based on PCA and LSTM to the field of wind power prediction [27].

III. LSTM-BASED ANOMALY DETECTION
In this section, LSTM-based multi-step prediction for non-linear dynamic system is presented in detail. Then, we propose an adaptive anomaly detection method and a local average algorithm with adaptive parameters.

In our proposed method, the first and last LSTM layers are followed by a dropout layer, which helps prevent overfitting. The last dropout layer is followed by a fully connected layer, as shown in Fig. 1. In the framework of the LSTM-based anomaly detection approach, the first segment is the data preprocessing procedure. After preprocessing, the mean value and threshold are calculated for the following parts. The raw data are transformed into normalized data. The second segment is the network architecture for multi-step prediction. It is composed of three kinds of network layers which include LSTM layer, dropout layer, and a fully connected layer. The seven-layer network outputs multi-step prediction results of the future time series values. In the third segment, the prediction results produced in the second segment are used as input and anomaly detection results are generated according to the local average algorithm with adaptive parameters.

A. LSTM-BASED MULTI-STEP PREDICTION
LSTM layer receives a time sequence of the same length N as input and outputs them to dropout layer to prevent overfitting. To collect samples for training, there are two different modes according to the requirement of prediction demand. As is shown in Fig. 2, there are two types of training modes according to the requirement of prediction demand. T refers to the length of the input time series and t means the prediction of the tth point after the current point.

In mode a, only the next point of current input time sequence is predicted in each prediction, as is illustrated in Fig. 2(a). A sample used for training is obtained by current state window. It contains the time sequence of fixed length and the value of the next timestep. By sliding over the long time series used for training, all samples are gathered into the training set. Mode b, as shown in Fig. 2(b), meets the requirement of muti-timestep prediction. For the sliding window in mode b, it equals to $T + t$ which contains the time sequence and a time interval between current value and the value to predict. In this mode, a training sample consists of
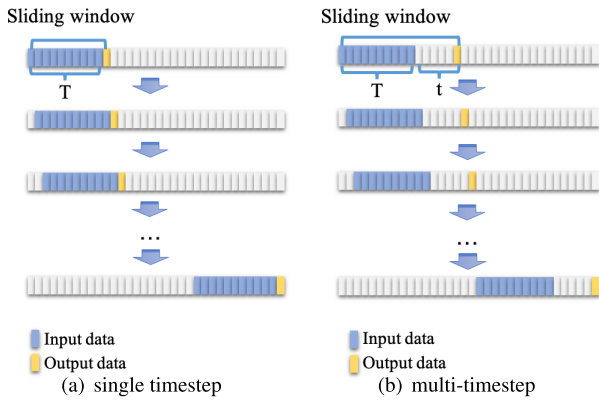
Sliding window

Sliding window

Input data
Output data
(a) single timestep

Input data
Output data
(b) multi-timestep

FIGURE 2. Two types of training modes.

time sequence of fixed length and the value of the tth timestep after the current value.

When all training samples are gathered into the training set, the training process begins. It randomly selects from the training set to remove correlations in the sequence and smooths the changes in data distribution.

Irrespective of different training modes, the Root Mean Square Error (RMSE) is utlized to evaluate the prediction performance of the proposed method,

$$ \text{RMSE} = \sqrt{\frac{1}{Q-T-t+1} \sum_{i=1}^{Q-T-t+1} (y_i - \hat{y}_i)^2}, \quad (1) $$

where y_i and $\hat{y}_i$ are the predicted value and the ground truth for timestep i, respectively, and Q is the total length of non-linear dynamical dataset used for training. Traditional models that be used for multi-timestep prediction include Auto Regressive Integrated Moving Average (ARIMA), eXtreme Gradient Boosting (XGBoost), Logistic Regression (LR), etc. Among them, ARIMA is used as our baseline model for comparison in this paper.

B. ADAPTIVE ANOMALY DETECTION

Another important module in the architecture of LSTM-based anomaly detection method is the following anomaly detection. An algorithm, named Local Average with Adaptive Parameters (LAAP), is put forward. This algorithm uses adaptive parameters to compute local numerical features including average, standard deviation and slope. The threshold for determining anomaly depends on those features of predicted timestep.

As the sliding window moves forward, the predicted time series can be obtained. The length of the predicted future timesteps is assumed to be d. The predicted time series can be denoted as $D : s_1, s_2, \ldots, s_d$. The window length and parameter adaptive rate are set to be W and α respectively. The value of W is usually determined by the pattern of most anomaly. The parameter adaptive rate can be adjusted according to how sensitive the detection system should be. It ranges from 0 to 1. If the parameter adaptive rate is relatively

high, the anomaly detection system is more sensitive to the upcoming anomaly. Otherwise, the system may reduce the probability for predicting an upcoming anomaly.

In each window length of W, the local average μ_i is computed as

$$ \mu_i = \frac{1}{W} \sum_{l=i-\frac{W}{2}}^{i+\frac{W}{2}} s_l, \quad (2) $$

and the local standard deviation σ_i is computed as

$$ \sigma_i = \sqrt{\frac{1}{W} \sum_{k=i-\frac{W}{2}}^{i+\frac{W}{2}} ((s_k - \mu_i)^2)}. \quad (3) $$

Then, a local slope k_i is computed to discriminate different kinds of potential anomalies. Generally, there are two kinds of potential anomalies in the non-linear dynamical system.

One kind of anomalies appears to be a local maximum while another kind of anomalies appears to be a local minimum. According to these two different situations, the result of anomaly detection y_i is obtained and output by algorithm LAAP. To summarize the above procedure of our proposed method, the pseudocode of LAAP is shown in Algorithm 1.

Algorithm 1 Local Average With Adaptive Parameters

Input:
the predicted time series, $D : s_1, s_2, \ldots, s_d$
window length, W
parameter adaptive rate, α

Process:

1: Compute the local average μ_i in a window of length W for every point in the predicted time series by (2)
2: Compute the local standard deviation σ_i in a window of length W for every point in the predicted time series by (3)
3: Compute the local slope k_i in a window of length W for every point in the predicted time series, $k_i = \frac{1}{W} \left(s \left(i + \frac{W}{2} \right) - s \left(i - \frac{W}{2} \right) \right)$
4: **if** $k(i) < 0$ **then**
5: $\quad y_i = \begin{cases} 1 & \text{if } s_i < \mu_i - \alpha \cdot \sigma_i \\ 0 & \text{otherwise} \end{cases}$
6: **else**
7: $\quad y_i = \begin{cases} 1 & \text{if } s_i > \mu_i + \alpha \cdot \sigma_i \\ 0 & \text{otherwise} \end{cases}$
8: **end if**

Output: y_i

IV. EXPERIMENTAL RESULTS AND ANALYSIS

In this section, experiments on the basis of non-linear dynamical data are carried out to demonstrate the effectiveness of our proposed method.

A. DATASET

To test and verify the LSTM-based multi-step prediction and the following anomaly detection, the data is generated by the temporal evolution of an incompressible Newtonian fluid in the plane Poiseuille (channel) geometry where the flow is driven by a constant volumetric flux, Q, at a Reynolds number of $R_e = 1800$. The x, y, and z coordinates are aligned in the streamwise, wall-normal, and spanwise directions, respectively. Periodic boundary conditions are imposed in the x and z direction with fundamental periods of L_x and L_z. No-slip boundary conditions are imposed at the top and bottom walls $y = \pm h$, where $h = L_y/2$ is the half-channel height. Such a boundary condition necessitates that streamwise velocity is zero at both walls. In this study, the so-called minimal flow unit methodology is used with a domain size of $L_x \times L_y \times L_z = 2\pi \times 2 \times \pi$ [28]. A numerical grid system is generated on $N_x \times N_y \times N_z$ meshes, where Fourier-Chebyshev-Fourier spectral spatial discretization is applied to all variables. With a mesh convergence study, the resolution used is $(N_x, N_y, N_z) = (48, 81, 48)$. The time step used for forward integration of the system is $\Delta t = 0.02$. Given these temporal and spatial resolutions, the corresponding errors are $O(10^{-6})$ and $O(10^{-4})$, respectively. For more details, the reader is referred to [29]–[31].

For this study, the one-dimensional data used for our experiments is the time series of the wall shear stress $\tau = \mu \partial u / \partial y$, where μ is the dynamic viscosity of the fluid and u is the streamwise velocity. The wall shear stress is a measure of the resistance of fluid experiences and is a qualitative measure of the state of nonlinear dynamical system.

A segment of the raw data is visualized in Fig. 3. The data fluctuates around a certain value and has a definite boundary in normal state. According to the value, time series can be divided into four zones: those below 90% of the mean value (zone 1), those between 90% of the mean value and the mean value (zone 2), those between the mean value and 110% of the mean value (zone 3), and those over 110% of the mean value (zone 4). Data in zone 4 should be recognized because there is a great probability of a turbulence. Usually, there is a high transition probability from zone 1 to zone 4, and anomalies are more likely to appear in zone 1 and zone 4. Hence, the main objective of the anomaly detection is to predict the data in zone 1 and zone 4. Accurate prediction of these anomalies can effectively prevent some potential accidents from happening.

B. RESULTS AND ANALYSIS

In this part, the experimental results are given and analyzed from two aspects: multi-step prediction and anomaly detection.

1) MULTI-STEP PREDICTION

The length of the input time sequence is 50 and the predicted timestep is set to be 4, 6, 8, 10 respectively. A segment of the test result consisting of 800 points is shown in Fig. 4.
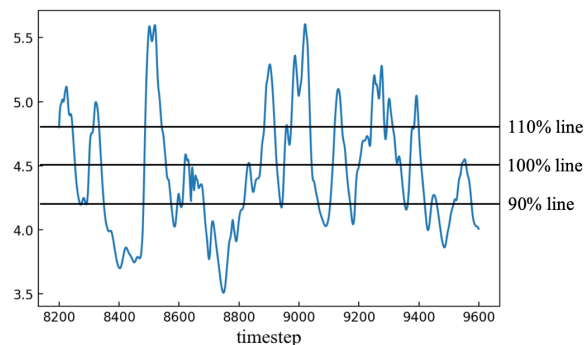
FIGURE 3. A segment of the raw data.

TABLE 1. Prediction accuracy of zone 1 and zone 4.

Timestep	Accuracy of zone 1	Accuracy of zone 4
1	0.9909	0.9991
2	0.9913	0.9932
3	0.9801	0.9949
4	0.9755	0.9910
5	0.9752	0.9850
6	0.9633	0.9807
7	0.9469	0.9850
8	0.9490	0.9743
9	0.9357	0.9739
10	0.9346	0.9713

The blue line refers to true data while the orange line refers to multi-step prediction result. As is shown in the four sub-figures in Fig. 4, the prediction error becomes larger and the fluctuation becomes more obvious when the predicted timestep increases.

Different LSTM network models are trained to predict the points of different step sizes. The whole data are split into two parts, i.e., training set and test set. The ratio of these two sets is 0.8. In this way, about 4000 samples are used for training and 1000 samples for testing. Table 1 shows the precision of zone 1 and zone 4 prediction. The mean value of the data set is 4.042179. The upper threshold is 4.446397 and the lower threshold is 3.637961. LSTM has achieved a high precision rate in both zone 1 and zone 4 prediction. Table 1 shows the prediction accuracy of zone 1 and zone 4 as the prediction timestep ranges from 1 to 10.

The prediction result of the LSTM-based method is compared with the ARIMA. When the predicted step size is larger than 2, LSTM performs better than ARIMA as the RMSE of LSTM predicted results is lower than that of the ARIMA. Furthermore, LSTM shows a stable growth, which means that different requirements of accuracy can be met by adjusting the predicted step size. The parameters for the ARIMA model *ar* and *ma* are both set to be 2 in our experiment. Then, the maximum likelihood estimation is used to fit our return rate to the ARIMA model. However, it should be noted that there are some irregular points in the ARIMA model. That is because the ARIMA model relies on the stability of temporal data and more stable data leads to better performance in prediction. However, the prediction of dynamical system contains one
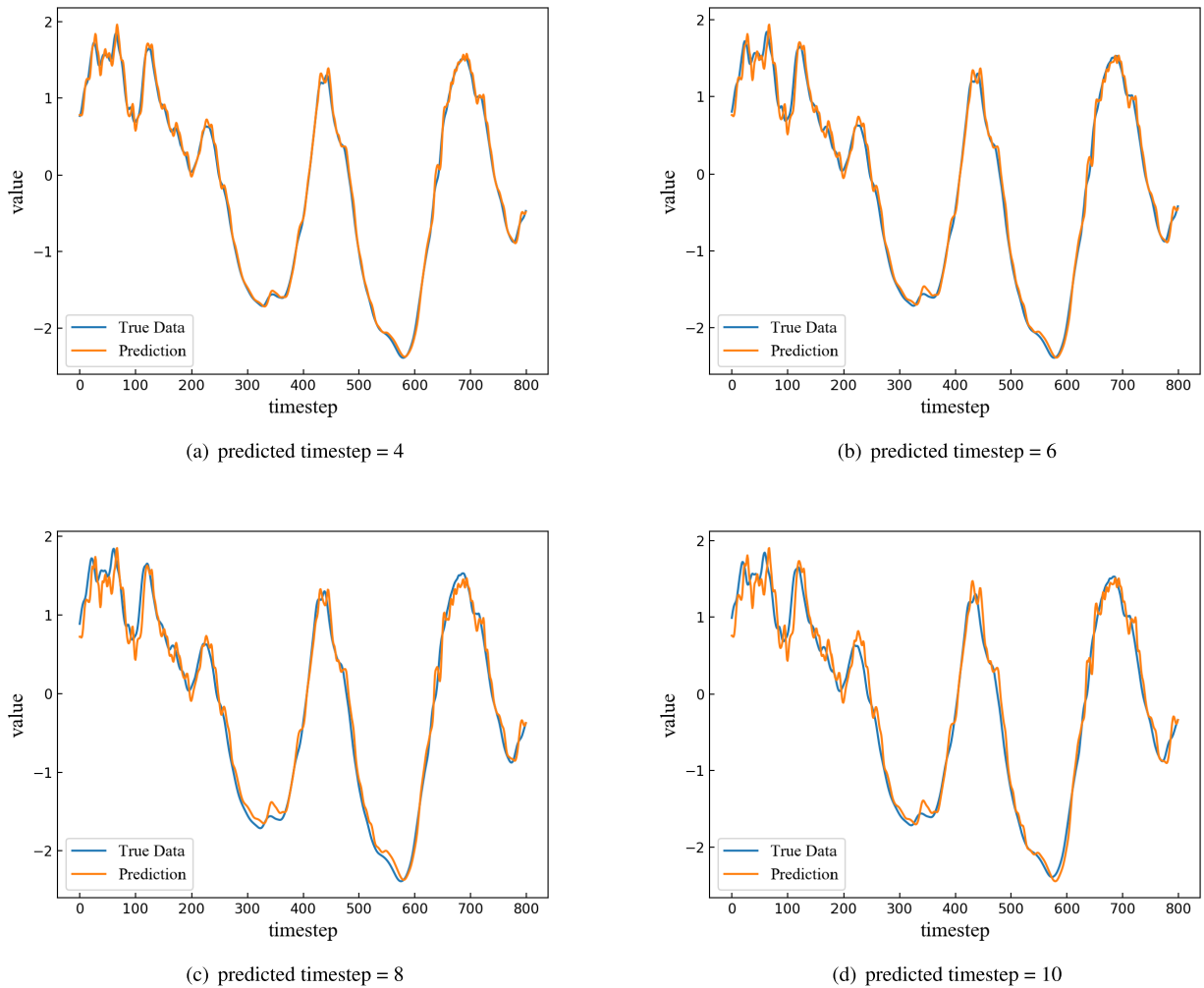
(a) predicted timestep = 4

(b) predicted timestep = 6

(c) predicted timestep = 8

(d) predicted timestep = 10

FIGURE 4. LSTM-based multi-step prediction result.

or more unstable factors. The ARIMA cannot capture such a change in time like LSTM.

2) ANOMALY DETECTION

We aim to raise alerts before anomalies occur. So, the anomaly detection results are acquired based on the multi-step prediction results. In order to prevent anomalies in advance, the anomaly detection model should be efficient and accurate. In comparison with absolute value-based anomaly detection, the accuracy of our proposed algorithm LAAP is evaluated by an error distribution function.

Fig. 6(a) and Fig. 6(b) illustrate the error distribution of absolute value-based anomaly detection algorithm and the LAAP algorithm, respectively. The horizontal axis represents the time interval error between the nearest true anomaly and predicted anomaly and the vertical axis refers to the total amount of anomalies of a certain predictive error. In Fig. 6(a), zero error occurs 14 times, and the time interval error ranges from around −175 to 75. While in Fig. 6(b), zero error occurs more than 20 times, and the time interval error ranges from around −130 to 5.
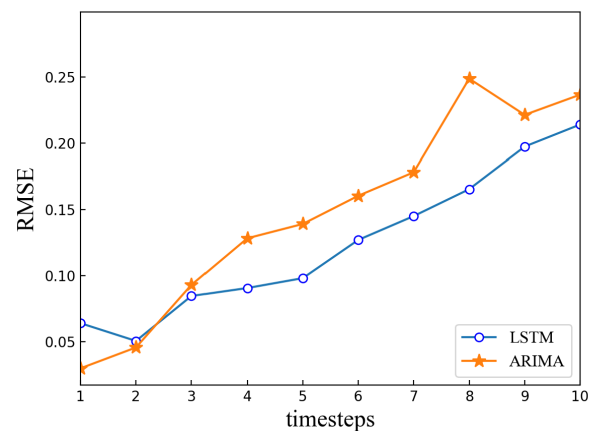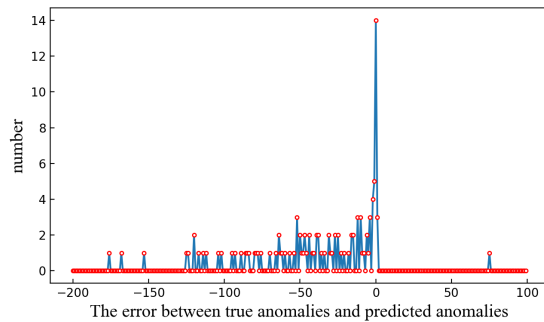

FIGURE 5. RMSE of different predicted step size.

According to the error distribution results, the frequency of zero error by LAAP algorithm is larger than that computed by the absolute value-based anomaly detection algorithm. Moreover, the error range of LAAP is smaller than the absolute value-based anomaly detection algorithm. Therefore,

(a) error distribution of absolute value-based anomaly detection algorithm
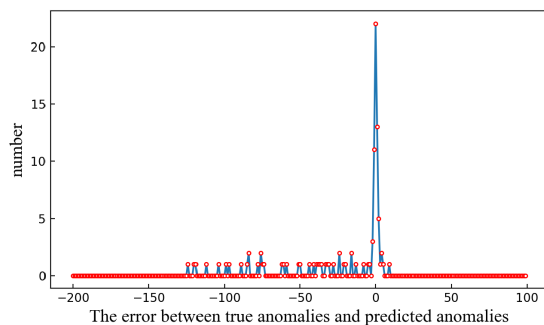
(b) error distribution of LAAP algorithm

FIGURE 6. Error distribution of anomaly detection algorithm.

it is more likely to accurately predict when an anomaly occurs by LAAP algorithm, and it can be concluded that LAAP performs better than the absolute value-based anomaly detection algorithm.

V. CONCLUSION

In this paper, we have proposed an LSTM-based anomaly detection method for non-linear dynamical system. A sliding window scheme is used to collect training samples for multi-step prediction. Then, an LAAP algorithm is developed to make anomaly detection based on the LSTM prediction result. Experiments have been conducted to evaluate the performance of the proposed methods. The results indicate that our proposed multi-step prediction has achieved lower RMSE and higher prediction accuracy compared with ARIMA on wall shear stress dataset, and the LAAP algorithm outperforms the absolute value-based anomaly detection algorithm.

REFERENCES

[1] P. Benner, S. Gugercin, and K. Willcox, "A survey of projection-based model reduction methods for parametric dynamical systems," *SIAM Rev.*, vol. 57, no. 4, pp. 483–531, Jan. 2015.

[2] R. Maulik, O. San, A. Rasheed, and P. Vedula, "Subgrid modelling for two-dimensional turbulence using neural networks," *J. Fluid Mech.*, vol. 858, pp. 122–144, Jan. 2019.

[3] C. Meneveau and J. Katz, "Scale-invariance and turbulence models for large-eddy simulation," *Annu. Rev. Fluid Mech.*, vol. 32, no. 1, pp. 1–32, Jan. 2000.

[4] C.-H. Moeng, "A Large-Eddy-Simulation model for the study of planetary boundary-layer turbulence," *J. Atmos. Sci.*, vol. 41, no. 13, pp. 2052–2062, Jul. 1984.

[5] F. Bassi, A. Crivellini, S. Rebay, and M. Savini, "Discontinuous Galerkin solution of the Reynolds-averaged Navier–Stokes and k-ω turbulence model equations," *Comput. Fluids*, vol. 34, nos. 4–5, pp. 507–540, 2005.

[6] M. Mortensen, H. P. Langtangen, and G. N. Wells, "A FEniCS-based programming framework for modeling turbulent flow by the Reynolds-averaged Navier–Stokes equations," *Adv. Water Resour.*, vol. 34, no. 9, pp. 1082–1101, Sep. 2011.

[7] N. Takeishi, Y. Kawahara, and T. Yairi, "Subspace dynamic mode decomposition for stochastic koopman analysis," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 96, no. 3, Sep. 2017, Art. no. 033310.

[8] A. Graves, M. Liwicki, S. Fernandez, R. Bertolami, H. Bunke, and J. Schmidhuber, "A novel connectionist system for unconstrained handwriting recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 5, pp. 855–868, May 2009.

[9] J. Yue-Hei Ng, M. Hausknecht, S. Vijayanarasimhan, O. Vinyals, R. Monga, and G. Toderici, "Beyond short snippets: Deep networks for video classification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 4694–4702.

[10] V. Chandola, A. Banerjee, and V. Kumar, "Anomaly detection: A survey," *ACM Comput. Surv.*, vol. 41, no. 3, p. 15, 2009.

[11] X. Gu and H. Wang, "Online anomaly prediction for robust cluster systems," in *Proc. IEEE 25th Int. Conf. Data Eng.*, Mar. 2009, pp. 1000–1011.

[12] A. Shameli Sendi, M. Dagenais, M. Jabbarifar, and M. Couture, "Real time intrusion prediction based on optimized alerts with hidden Markov model," *J. Netw.*, vol. 7, no. 2, p. 311, Feb. 2012.

[13] A. A. Paulo and L. S. Pereira, "Prediction of SPI drought class transitions using Markov chains," *Water Resour. Manage.*, vol. 21, no. 10, pp. 1813–1827, Sep. 2007.

[14] X. Zhou, S. Li, and Z. Ye, "A novel system anomaly prediction system based on belief Markov model and ensemble classification," *Math. Problems Eng.*, vol. 2013, pp. 1–10, Sep. 2013.

[15] D. Hong, D. Zhao, and Y. Zhang, "The entropy and PCA based anomaly prediction in data streams," *Procedia Comput. Sci.*, vol. 96, pp. 139–146, Jan. 2016.

[16] S. Huang, C. Fung, K. Wang, P. Pei, Z. Luan, and D. Qian, "Using recurrent neural networks toward black-box system anomaly prediction," in *Proc. IEEE/ACM 24th Int. Symp. Qual. Service (IWQoS)*, Jun. 2016, pp. 1–10.

[17] J. Peng, Z. Huang, and J. Cheng, "A deep recurrent network for Web server performance prediction," in *Proc. IEEE 2nd Int. Conf. Data Sci. Cyberspace (DSC)*, Jun. 2017, pp. 500–504.

[18] E. Zio, M. Broggi, L. Golea, and N. Pedroni, "Failure and reliability predictions by infinite impulse response locally recurrent neural networks," *Chem. Eng. Trans.*, vol. 26, pp. 1–6, Jun. 2012.

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

[20] K. Greff, R. K. Srivastava, J. Koutnik, B. R. Steunebrink, and J. Schmidhuber, "LSTM: A search space odyssey," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 10, pp. 2222–2232, Oct. 2017.

[21] P. Malhotra, L. Vig, G. Shroff, and P. Agarwal, "Long short term memory networks for anomaly detection in time series," in *Proc. Conf., 23rd Eur. Symp. Artif. Neural Netw., Comput. Intell. Mach. Learn. (ESANN)*. Ottignies-Louvain-la-Neuve, Belgium: Presses univ. de Louvain, 2015, p. 89.

[22] Y. Duan, Y. Lv, and F.-Y. Wang, "Travel time prediction with LSTM neural network," in *Proc. IEEE 19th Int. Conf. Intell. Transp. Syst. (ITSC)*, Nov. 2016, pp. 1053–1058.

[23] X. Jia, A. Khandelwal, G. Nayak, J. Gerber, K. Carlson, P. West, and V. Kumar, "Incremental dual-memory LSTM in land cover prediction," in *Proc. 23rd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining (KDD)*. New York, NY, USA: ACM, 2017, pp. 867–876.

[24] S. Yuan, X. Wu, and Y. Xiang, "Incorporating pre-training in long short-term memory networks for tweet classification," *Social Netw. Anal. Mining*, vol. 8, no. 1, p. 52, Dec. 2018.

[25] K. Chen, Y. Zhou, and F. Dai, "A LSTM-based method for stock returns prediction: A case study of China stock market," in *Proc. IEEE Int. Conf. Big Data (Big Data)*, Oct. 2015, pp. 2823–2824.

[26] F. Altche and A. de La Fortelle, "An LSTM network for highway trajectory prediction," in *Proc. IEEE 20th Int. Conf. Intell. Transp. Syst. (ITSC)*, Oct. 2017, pp. 353–359.

[27] Q. Xiaoyun, K. Xiaoning, Z. Chao, J. Shuai, and M. Xiuda, "Short-term prediction of wind power based on deep long short-term memory," in *Proc. IEEE PES Asia–Pacific Power Energy Eng. Conf. (APPEEC)*, Oct. 2016, pp. 1148–1152.

[28] J. Jiménez and P. Moin, "The minimal flow unit in near-wall turbulence," *J. Fluid Mech.*, vol. 225, pp. 213–240, Apr. 1991.

[29] J. S. Park and M. D. Graham, "Exact coherent states and connections to turbulent dynamics in minimal channel flow," *J. Fluid Mech.*, vol. 782, pp. 430–454, Nov. 2015.

[30] J. S. Park, A. Shekar, and M. D. Graham, "Bursting and critical layer frequencies in minimal turbulent dynamics and connections to exact coherent states," *Phys. Rev. Fluids*, vol. 3, no. 1, Jan. 2018, Art. no. 014611.

[31] R. D. Whalley, J. S. Park, A. Kushwaha, D. J. C. Dennis, M. D. Graham, and R. J. Poole, "Low-drag events in transitional wall-bounded turbulence," *Phys. Rev. Fluids*, vol. 2, no. 3, Mar. 2017, Art. no. 034602.

KAN ZHENG (Senior Member, IEEE) received the B.S., M.S., and Ph.D. degrees from the Beijing University of Posts and Telecommunications (BUPT), China, in 1996, 2000, and 2005, respectively. He is currently a Full Professor with BUPT. He has rich experience in research and standardization of new emerging technologies. He has authored over 200 journal articles and conference papers in the field of wireless communications, vehicular networks, the IoT, security, and so on. He holds editorial board positions with several journals. He has organized several special issues in the journals, including the IEEE COMMUNICATIONS SURVEYS AND TUTORIALS, the *IEEE Communications Magazine*, and the IEEE SYSTEMS JOURNAL. He has also served in the organizing/TPC committees for more than ten conferences.

YUE TAN (Graduate Student Member, IEEE) received the B.S. degree from the Beijing University of Posts and Telecommunications, Beijing, China, in 2017, where she is currently pursuing the master's degree. Her current research interests include deep reinforcement learning and its applications in the Internet of Things and smart grid.

ETHAN A. DAVIS received the B.S. degree in mechanical engineering and the M.S. degree in mechanical engineering and applied mechanics from the University of Nebraska-Lincoln, in 2014 and 2017, respectively, and the Ph.D. degree in mechanical engineering and applied mechanics. Since 2018, he has been with the Park Research Group, University of Nebraska-Lincoln. His research interests include turbulence dynamics, the application of flow control methods for increased energy savings, and the application of machine learning in turbulent flow prediction and control. He is a member of the American Physics Society (APS) and the American Society of Mechanical Engineers (ASME). He received the Dean's Fellowship from the University of Nebraska-Lincoln and the NASA Nebraska Space Grant Fellowship.

CHUNJING HU received the B.S., M.S., and Ph.D. degrees from the Beijing University of Posts and Telecommunications (BUPT), in 1991, 1994, and 2007, respectively. She is currently an Associate Professor with the Key Laboratory of Universal Wireless Communications, Ministry of Education, BUPT. She has published over 60 journal articles and conference papers. Her research interests include speech recognition, machine translation, and the application of wireless communication techniques.

JAE SUNG PARK received the B.S. degree in mechanical engineering from Hanyang University, Seoul, South Korea, and the M.S. and Ph.D. degrees in mechanical engineering from the University of Illinois at Urbana–Champaign, in 2008 and 2012, respectively. He was trained as a Postdoctoral Associate with the Department of Chemical and Biological Engineering, University of Wisconsin-Madison. He is currently an Assistant Professor with the Department of Mechanical and Materials Engineering, University of Nebraska-Lincoln. His research interests include wide range of fluid mechanics, covering from low to high Reynolds number flows, including complex fluids, electrokinetics, biofluids, and turbulence. He also studies thermo-fluids systems such as heat exchangers. His research has been funded by the National Science Foundation, NASA, and Nebraska Research Initiatives. He is a member of the American Physics Society (APS), the American Society of Mechanical Engineers (ASME), and the Korean American Scientists and Engineers Association (KSEA). He was awarded a 2020 recipient of the College of Engineering Henry Y. Kleinkauf Family Distinguished New Faculty Teaching Award at the University of Nebraska-Lincoln.

KUAN ZHANG (Member, IEEE) received the B.Sc. degree in communication engineering and the M.Sc. degree in computer applied technology from Northeastern University, Shenyang, China, in 2009 and 2011, respectively, and the Ph.D. degree in electrical and computer engineering from the University of Waterloo, Waterloo, ON, Canada, in 2016. He was a Postdoctoral Fellow with the Broadband Communications Research Group, Department of Electrical and Computer Engineering, University of Waterloo, from 2016 to 2017. He has been an Assistant Professor with the Department of Electrical and Computer Engineering, University of Nebraska-Lincoln, Omaha, NE, USA, since 2017. His current research interests include security and privacy for mobile social networks, e-healthcare systems, cloud/edge computing, and cyber physical systems. He was a recipient of the Best Paper Award in IEEE WCNC 2013, Securecomm 2016, and BigDataSE 2019.

● ● ●