

Research Article

Prediction for Chaotic Time Series-Based AE-CNN and Transfer Learning

Baogui Xin  and Wei Peng

College of Economics and Management, Shandong University of Science and Technology, Qingdao 266590, China

Correspondence should be addressed to Baogui Xin; xin@sdust.edu.cn

Received 31 July 2020; Revised 28 August 2020; Accepted 10 September 2020; Published 16 September 2020

Academic Editor: Min Xia

Copyright © 2020 Baogui Xin and Wei Peng. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

It has been a hot and challenging topic to predict the chaotic time series in the medium-to-long term. We combine autoencoders and convolutional neural networks (AE-CNN) to capture the intrinsic certainty of chaotic time series. We utilize the transfer learning (TL) theory to improve the prediction performance in medium-to-long term. Thus, we develop a prediction scheme for chaotic time series-based AE-CNN and TL named AE-CNN-TL. Our experimental results show that the proposed AE-CNN-TL has much better prediction performance than any one of the following: AE-CNN, ARMA, and LSTM.

1. Introduction

Chaos is the unity of extrinsic randomness and intrinsic certainty, which widely exists in social fields and natural systems. However, in many cases, we cannot represent chaotic systems by establishing analytical mathematical models but by observing the time series [1]. With the development of big data acquisition and research technology, we can obtain large-scale chaotic time series from a wide range of social fields and natural systems, which highlight the necessity of research on the prediction of chaotic time series. Of course, the prediction of chaotic systems has always been a hot and challenging topic in many fields. However, the acquisition of chaotic time series data usually accompanies errors, which make the prediction of chaotic time series more difficult.

The most significant characteristic of a chaotic system is its initial sensitivity; that is, a slight change in input will produce random and unpredictable difference after long enough evolution, which determines the difficulty in medium-to-long-term prediction for a chaotic system. Similarly, the error in chaotic time series may be amplified to an unacceptable degree due to its initial sensitivity, which will lead to the failure of prediction. However, the orbit of the chaotic system also has intrinsic certainty such as self-

similarity, which indicates that a chaotic system can be predicted to a certain extent, even in the medium-to-long term by using fractal geometry and other related theories. But the self-similarity and other intrinsic certainty of a chaotic system are very difficult to be measured accurately. So far, there is no commonly accepted theory that can be employed to accurately measure the intrinsic certainty such as self-similarity and realize the medium-to-long-term prediction of a chaotic system. Inspired by Xia et al. [2], we try to overcome the following difficulties to effectively predict the medium-to-long-term time series: (1) how to make full use of the intrinsic certainty of chaotic time series? It needs us to propose or borrow a universal approximation framework [3], which has a powerful representation of deep learning and can extract high-order features from input information for medium-to-long-term prediction of chaotic time series; (2) how to extract intrinsic high-order features from a chaotic time series and transfer them to its coupled time series to strengthen its prediction performance.

To solve abovementioned difficulties, we will mainly implement the following contributions:

- (1) Employ a convolutional neural network (CNN) to extract high-order features of a chaotic time series for medium-to-long-term prediction. The CNN has

a strong representation learning ability, so we can use it to classify the input information according to its hierarchical structure and extract high-order features from the input information.

- (2) Utilize the transfer learning (TL) theory to improve the medium-to-long-term prediction performance. We extract the intrinsic high-order features from the chaotic time series and use the transfer learning theory to overlay them with their high-order features to enhance its prediction performance.

The remainder of this article is organized as follows: in Section 2, we review the related work; in Section 3, we describe the proposed the AE-CNN-TL in detail; in Section 4, we implement some experiments and discuss the results; and in Section 5, we draw a conclusion and propose some future works.

2. Related Work

2.1. Classification of Prediction Methods for Chaotic Time Series. According to the step number of prediction, we can divide the prediction methods into the single-step prediction [4] and multistep prediction [5, 6]. In the single-step prediction, the model can be only used to predict the sequence values of the next step. In the multistep prediction, the model can be employed to predict the sequence values of the multistep, which sets higher requirements. There are two strategies of the multistep prediction: direct approach and iterative approach. Based on the direct approach, the prediction of chaotic time series can complete long-term prediction once by constructing a function mapping between input and multistep output. The idea of the direct approach is simple and easy to implement, but the direct approach needs more complex function mapping structure, which is a significant challenge. The iterative approach is to predict the time series value after only one step at a time and then take the predicted value as known and finally iterate to predict the next step. The iterative approach has low requirements for the model, but in the iterative process, the prediction error is also input into the model, which results in the error accumulation and the prediction accuracy decrease in medium-to-long term. The proposed AE-CNN-TL constructed in this article belongs to the direct approach.

2.2. Prediction of Chaotic Time Series Based on Deep Learning of Direct Strategies. Taking accurate multistep predictions with a novel recurrent neural network (RNN), Min et al. [5] constructed a new scheme to implement long-term prediction. Yi et al. [7] introduced a deep neural network-(DNN-) based method to predict the air quality in medium term. Kuo and Huang [8] presented a DNN model (EPNet), which combines CNN and LSTM to predict the electricity price in medium term. Mujeeb et al. [9] built a deep LSTM model for short- and medium-term prediction. Shen et al. [10] set up a prediction model for time series named SeriesNet, which includes an LSTM and a dilated causal convolution network. Mudassir et al. [11] established the high-performance machine learning- (ANN, SANN,

SVM, and LSTM) based classification and regression models to predict Bitcoin price in short-to-medium term. Hussain et al. [12] put forward a one-dimensional convolutional neural network (1D-CNN) and extreme learning machine (ELM) to predict the one-step-ahead stream-flow in short and medium terms. Chimmula and Zhang [13] constructed an LSTM network to predict COVID-19 cases. Livieris et al. [14] designed a CNN-LSTM model to predict gold prices. Zhang and Dong [15] provided a CRNN model consisting of CNN and RNN to predict temperature changes from historical data. Li [16] proposed the KLS method fusing Kalman filter, LSTM, and SVM to forecast the carbon emission. Combining multikernel convolutional layers and convolutional LSTM layers, Asadi and Regan [17] propose a deep learning framework for spatiotemporal forecasting problems.

3. Proposal

As we know, we can employ an autoencoder (AE) to learn a representation (encoding) for a time series by training the network to ignore signal "noise," i.e., denoising. Considering CNN's advantages of shift invariant or space invariant, the CNN is most commonly applied to classify images, cluster images by similarity, and perform object recognition within scenes. In this article, we will employ CNN to extract the intrinsic high-order features including similarity from the chaotic time series to implement predictions. By combining the advantages of AE and CNN, we can use AE-CNN to predict a chaotic time series, as shown in Figure 1.

If there are a chaotic time series Y associated with X , we can extract the high-order intrinsic features from Y and utilize the transfer learning theory to overlay with the high-order intrinsic features from X to enhance the prediction performance. Thus, we can develop a prediction scheme for chaotic time series of medium-to-long-term-based AE-CNN and transfer learning, named AE-CNN-TL, as shown in Figure 2.

We assume that both time series X and Y with length T , where $X = (x_1, x_2, \dots, x_N)^T \in \mathbb{R}^{N \times 1}$ and $Y = (y_1, y_2, \dots, y_N)^T \in \mathbb{R}^{N \times 1}$.

From Figure 2, AE-CNN-TL consists of four processes: denoising, feature extracting, transferring learning, and forecasting. At the denoising stage, a typical AE is firstly used for data filtering so as to remove the noises from the chaotic time series. At the feature extracting stage, the denoised chaotic time series X and original time series Y would be taken as the inputs of CNNs to extract deep representations. On the transferring learning stage, the deep representations from Y can be transferred to them from X and overlapped together. At the forecasting stage, we would obtain the prediction of X through a fully connected layer.

The optimization of AE-CNN-TL is to minimize the reconstruction error of AE and the training error of the whole model. At the denoising stage, the output of AE is an approximate copy of input. Therefore, we have to minimize the reconstruction error between input and output, which could maintain the direct significance of temperature. To obtain a sound model, it is necessary to minimize the

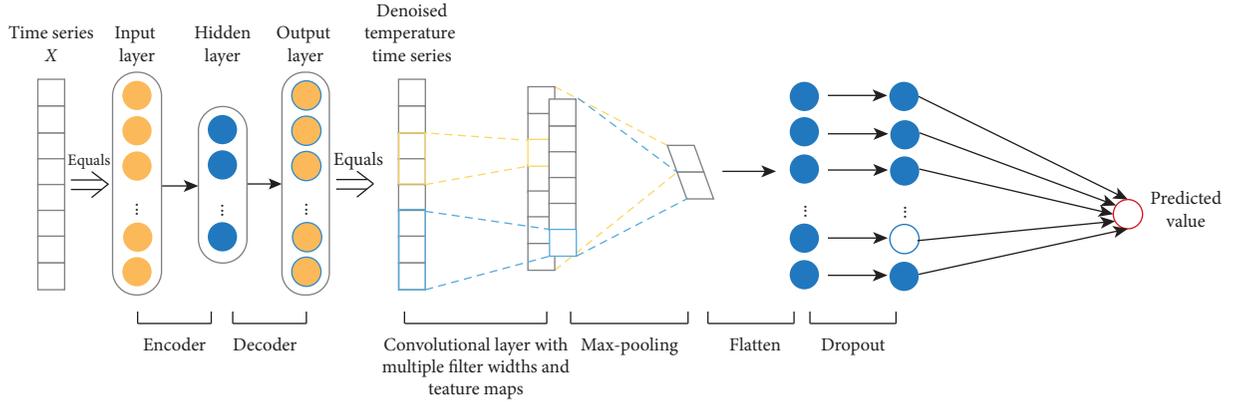


FIGURE 1: The architecture of AE-CNN.

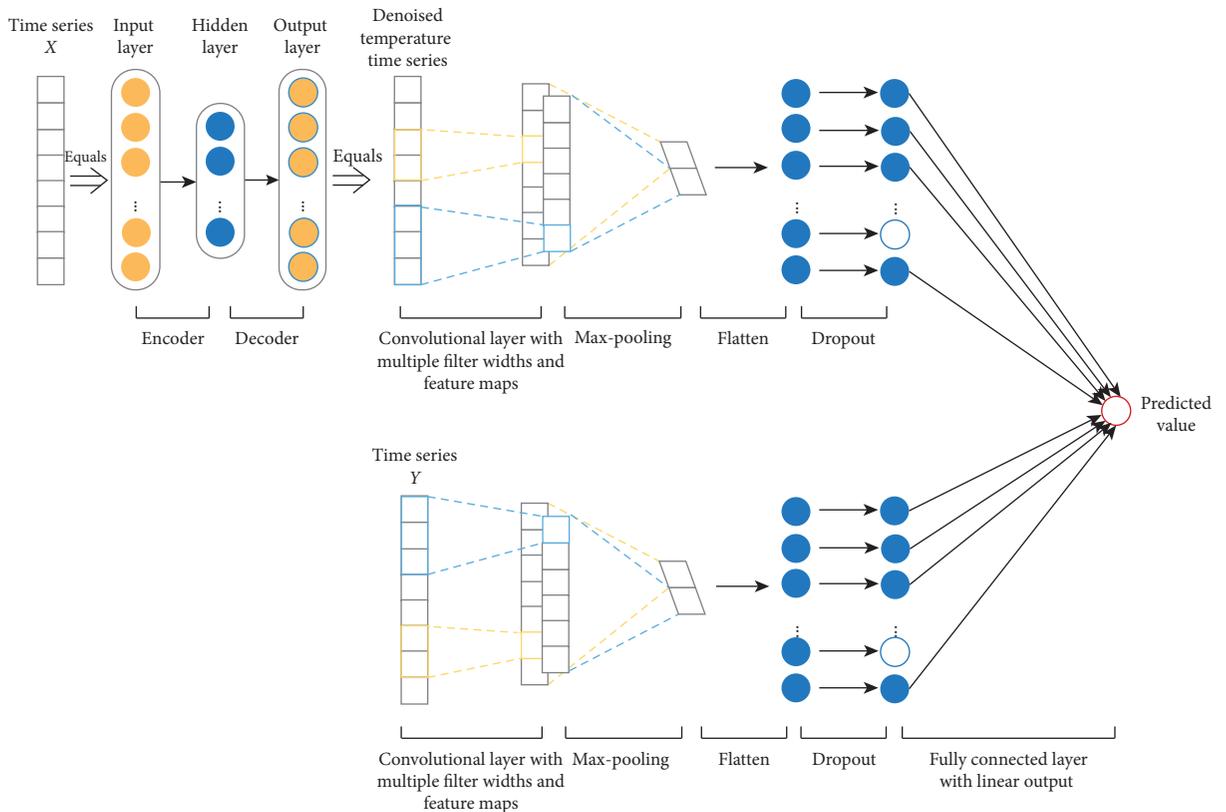


FIGURE 2: The architecture of AE-CNN-TL.

training error of the whole model. We formulate the objective functions as

$$e_{AE} = \min \sum \|X - g(W' \times f(WX + b) + b')\|^2, \quad (1)$$

where $f(\cdot)$ and $g(\cdot)$ are activation functions, W and W' are weights, and b and b' are biases.

It is necessary to obtain a sound model to minimize the training error of the whole model. We formulate the objective function of the whole model as

$$e_{AE-CNN-TL} = \min \sum \|x_t - \hat{x}_t\|^2, \quad (2)$$

where \hat{x}_t denotes the predicted value.

4. Experiments and Results

4.1. Data. In this article, we use the Beijing PM2.5 dataset (<https://archive.ics.uci.edu/ml/datasets/Beijing+PM2.5+Data>).

The hourly data span the period between January 1, 2010, and December 31, 2014. Each series in this dataset contains 43824 data points. After removing missing values, 41757 data points were finally used in the experiment. The time series we used is PM2.5 concentration ($\mu\text{g}/\text{m}^3$) and temperature ($^{\circ}\text{C}$). The descriptive statistics for PM2.5 concentration and temperature is shown in Table 1.

To obtain a desirable model, we divide the experimental data into three parts: 70% training dataset, 10% validation

TABLE 1: Descriptive statistics for PM2.5 concentration and temperature.

	PM2.5 concentration	Temperature
Mean	98.61	12.40
Median	72.00	14.00
Standard deviation	92.05	12.18
Sample variance	8,473.27	148.24
Kurtosis	7.77	1.88
Skewness	1.80	-0.14
Range	994.00	61.00
Minimum	0.00	-19.00
Maximum	994.00	42.00
Count	41,757	41,757

dataset, and 20% test dataset. The training dataset is to reach a sound model, the validation dataset is to further determine the parameters of the whole network, and the test dataset is to test the generalization ability of the model.

4.2. Evaluation Index. The root mean square error (RMSE) is a widely used evaluation index for continuous type forecasting. Generally speaking, the smaller the RMSE value is, the better its prediction performance is. The RMSE is defined as follows:

$$\text{RMSE} = \sqrt{\frac{\sum_{t=1}^{N_1} (x_t - \hat{x}_t)^2}{|N_1|}}, \quad (3)$$

where N_1 denotes the length of test dataset.

4.3. 0-1 Test for Chaos in the Beijing PM2.5 Dataset. According to the 0-1 test algorithm for chaos, we sample $M = 4,176$ data points from the Beijing PM2.5 dataset after taking every 10th data point and set a random number c drawn from a uniform distribution on $(\pi/5, 4\pi/5)$ satisfying the following translation equations:

$$\begin{cases} p(n) = \sum_{j=1}^n X(j)\cos(jc), \\ s(n) = \sum_{j=1}^n X(j)\sin(jc), \end{cases} \quad (4)$$

where X denotes the time series of PM2.5 or temperature.

Then, we calculate the following test value of chaos:

$$k = \text{median} \left[\frac{\text{cov}(\phi, \varphi)}{\sqrt{\text{cov}(\phi, \phi)\text{cov}(\varphi, \varphi)}} \right], \quad (5)$$

where $\phi = (1, 2, \dots, n_{\text{cut}})$, $\varphi = (D_c(1), 2D_c(2), \dots, D_c(n_{\text{cut}}))$, and the covariance is written with vectors ϕ and φ of length r as follows:

$$\text{cov}(\phi, \varphi) = \frac{1}{r} \sum_{j=1}^r \left[\phi(j) - \frac{1}{r} \sum_{j=1}^r \phi(k) \right] \left[\varphi(j) - \frac{1}{r} \sum_{j=1}^r \varphi(j) \right], \quad (6)$$

$$\begin{aligned} D_c(n) = \lim_{M \rightarrow \infty} \frac{1}{M} \sum_{j=1}^M \left[(p_c(j+n) - p_c(j))^2 \right. \\ \left. + (s_c(j+n) - s_c(j))^2 \right] \\ - \frac{1 - \cos nc}{1 - \cos c} \left[\lim_{N \rightarrow \infty} \frac{1}{M} \sum_{k=1}^M X(j) \right]^2. \end{aligned} \quad (7)$$

Figure 3 shows plots of the test value k vs. random number c for the time series of temperature. $k \approx 1$ implies that there is a chaotic attractor in the time series of temperature.

Figure 4 shows Brownian-like trajectories and denotes there is also a chaotic attractor in the time series of temperature. That is to say, the time series of temperature is chaotic.

Similar to Figure 3, Figure 5 also shows plots of the test value k vs. random number c for the PM2.5 concentration time series. $k \approx 1$ also implies there is a chaotic attractor in the time series of PM2.5 concentration.

Similar to Figure 4, Figure 6 also shows Brownian-like trajectories and denotes there is also a chaotic attractor in the time series of PM2.5 concentration. That is to say, the time series of PM2.5 concentration is chaotic.

Figure 7 demonstrates the phase portrait of temperature vs. PM2.5 concentration, in which Brownian-like (unbounded) trajectories imply there exists a chaotic attractor.

In short, Figures 3–7 prove that there exist chaotic attractors in the Beijing PM2.5 dataset.

4.4. Experimental Results. In this part, we test the prediction performance of the proposed AE-CNN-TL for chaotic time series.

To compare the prediction performance of AE-CNN and AE-CNN-TL, we firstly employ AE-CNN to predict PM2.5 concentration without TL from temperature, as shown in Figure 8. Then, we utilize AE-CNN-TL to predict PM2.5 concentration, as shown in Figure 9. In the same way, we firstly employ AE-CNN to predict temperature without TL from PM2.5 concentration, as shown in Figure 10. Then, we utilize AE-CNN-TL to predict temperature, as shown in Figure 11. Owing to the continuous value prediction, we employ RMSE to measure the prediction performance. The smaller the RMSE value is, the better the prediction

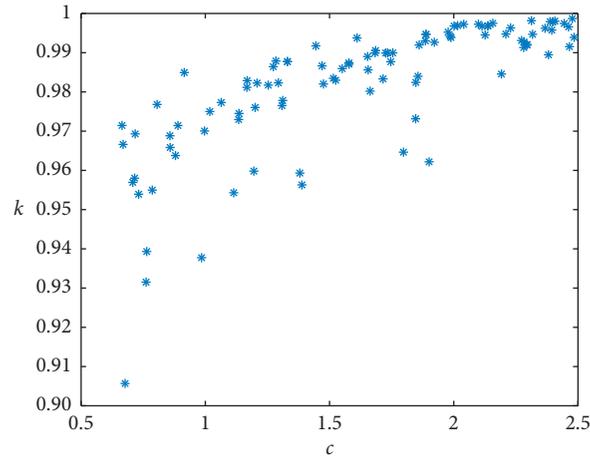


FIGURE 3: Plots of K vs. c for temperature.

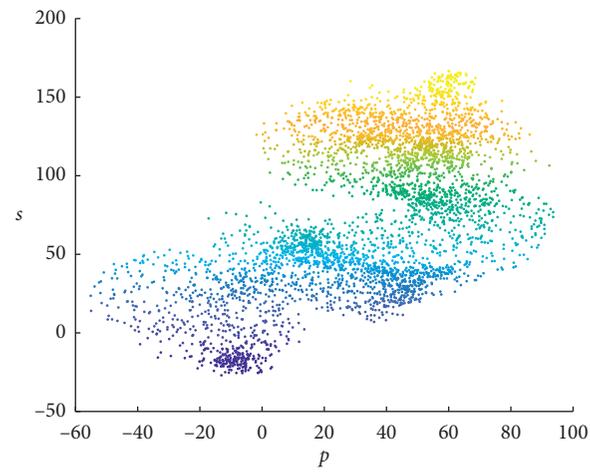


FIGURE 4: Plots in the new coordinate p - s plane for temperature.

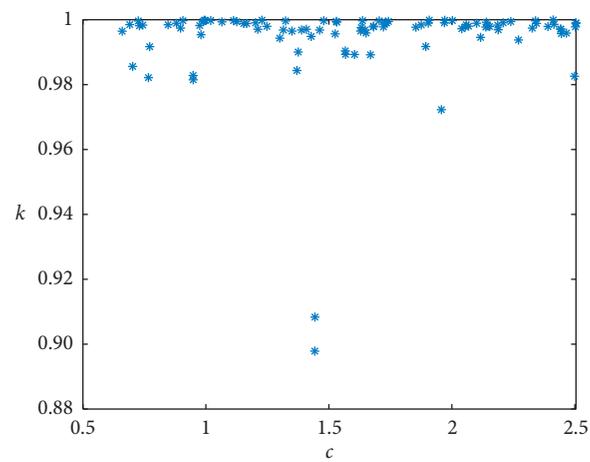


FIGURE 5: K vs. c for PM2.5 concentration.

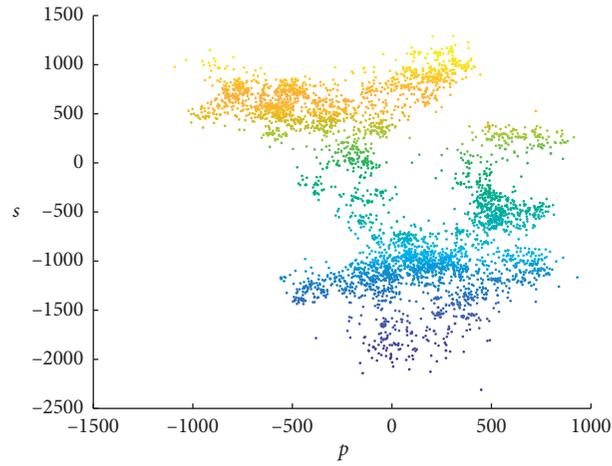


FIGURE 6: Phase in p - s plane for PM2.5 concentration.

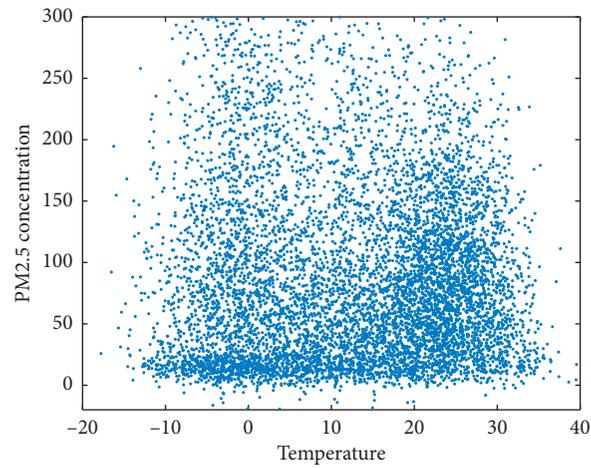


FIGURE 7: Phase portrait of temperature vs. PM2.5 concentration.

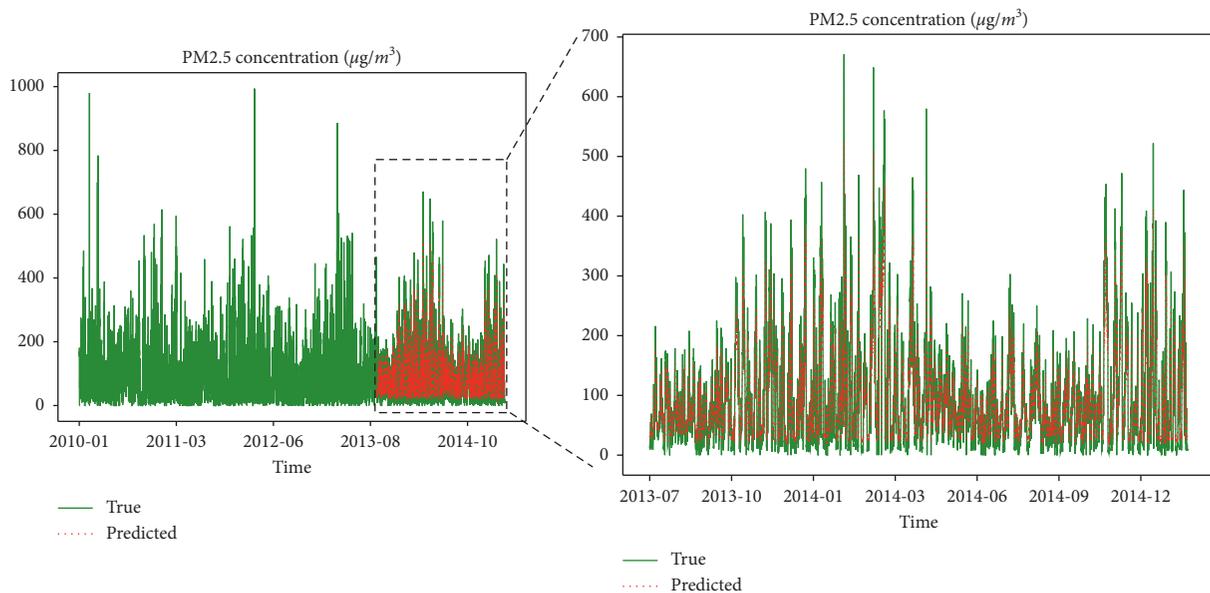


FIGURE 8: PM2.5 concentration prediction without TL from temperature.

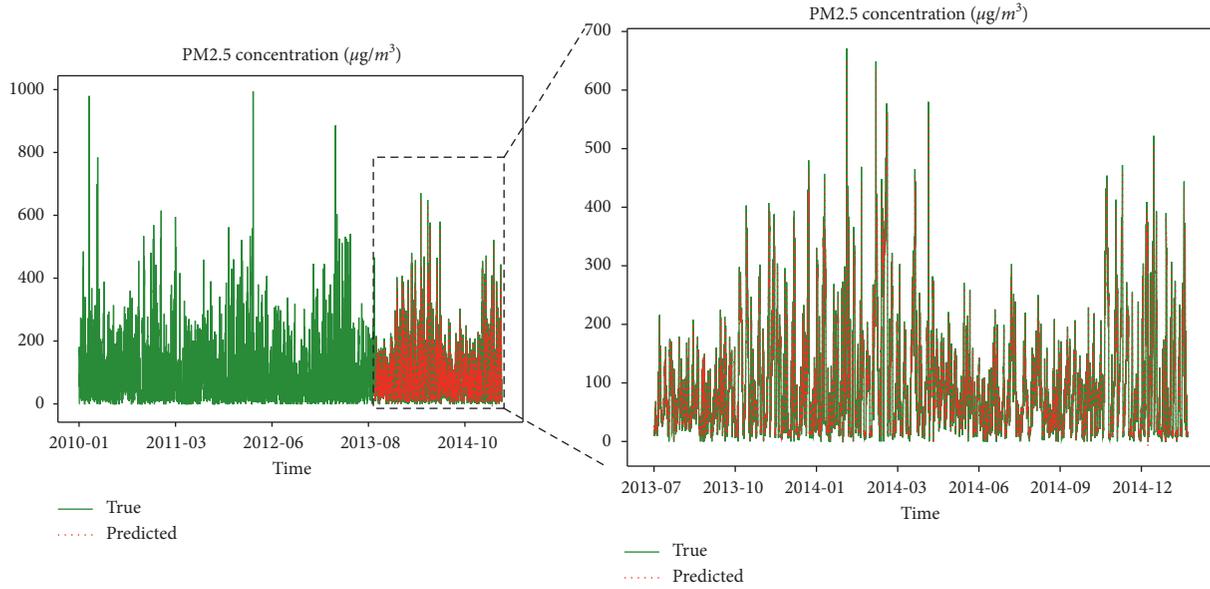


FIGURE 9: PM2.5 concentration prediction with TL from temperature.

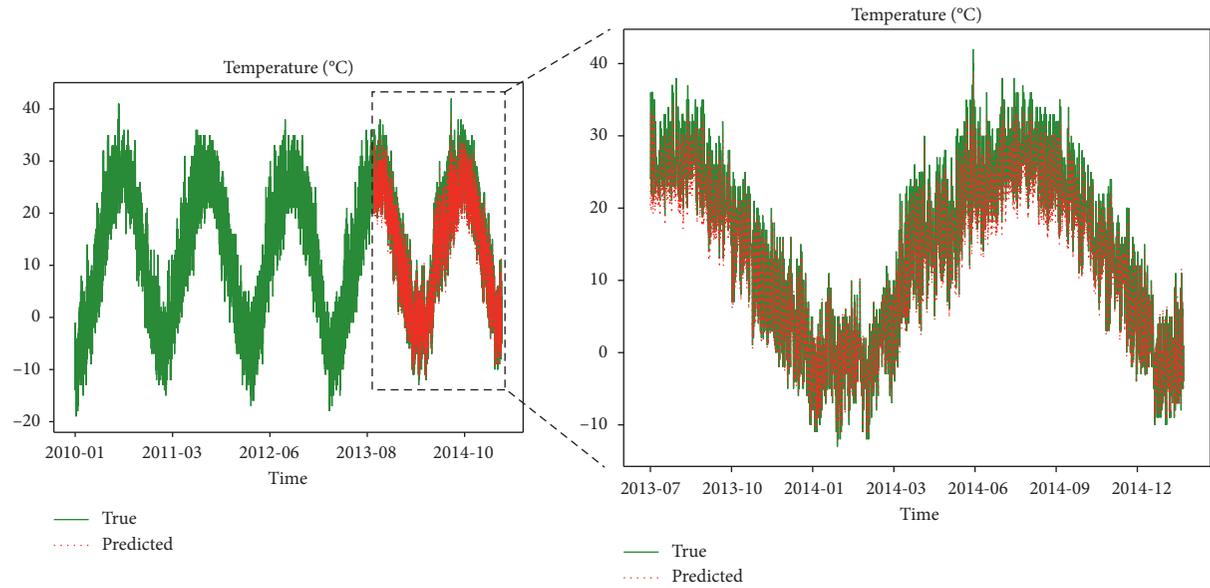


FIGURE 10: Temperature prediction without TL from PM2.5 concentration.

performance is. And these corresponding RMSEs of prediction are shown in Table 2.

From Table 2, we can catch that the prediction accuracy of PM2.5 concentration is enhanced to a large extent with taking TL from temperature. And the prediction accuracy

of temperature is also improved significantly with taking TL from PM2.5 concentration. Therefore, we can conclude that AE-CNN-TL is much better in the prediction performance than any one of the following: AE-CNN, ARMA, and LSTM.

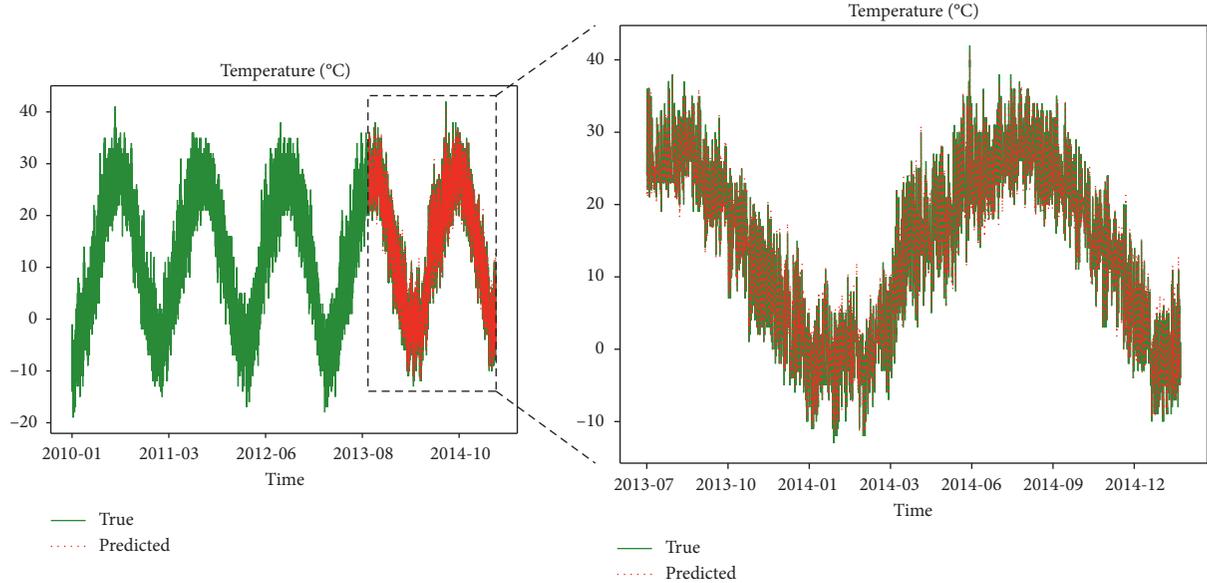


FIGURE 11: Temperature prediction with TL from PM2.5 concentration.

TABLE 2: RMSEs of prediction.

Prediction task	Prediction approach	RMSE
PM2.5 concentration prediction	ARMA	79.21
	LSTM	43.50
	AE-CNN	89.55
	AE-CNN-TL	21.53
Temperature prediction	ARMA	72.47
	LSTM	37.68
	AE-CNN	12.39
	AE-CNN-TL	1.61

5. Conclusions

In this article, we develop a prediction scheme for chaotic time series of medium-to-long-term-based AE-CNN and TL, named AE-CNN-TL, which has much better prediction performance than any one of the following: AE-CNN, ARMA, and LSTM. This prediction scheme can be used in chaotic time series of medium-to-long term of many natural fields and social systems in the real world.

In fact, AE-CNN-TL can also be used to reveal the dynamic link between PM2.5 concentration and temperature. What is more, AE-CNN-TL can be used to explore the Granger causality [18] between PM2.5 concentration and predict PM2.5 concentration and temperature with a good performance. By means of experiments, we find a bidirectional Granger causality between PM2.5 concentration and temperature, which is consist with previous researches.

We know that the Beijing PM2.5 dataset originates from the nature field. In fact, we can try to employ the proposed AE-CNN-TL to predict a chaotic time series of medium-to-long term which comes from some artificial systems, such as the game system [19–21] and the financial system [22].

Data Availability

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation, to any qualified researcher.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this article.

Acknowledgments

This work was supported by the Natural Science Foundation of Shandong Province (Grant no. ZR2016FM26) and the National Social Science Foundation of China (Grant no. 16FJY008).

References

- [1] J. Ma, Y. Chen, and B. Xin, "Study on prediction methods for dynamic systems of nonlinear chaotic time series," *Applied Mathematics and Mechanics*, vol. 25, pp. 605–611, 2004.
- [2] M. Xia, W. A. Liu, K. Wang, X. Zhang, and Y. Xu, "Non-intrusive load disaggregation based on deep dilated residual network," *Electric Power Systems Research*, vol. 170, pp. 277–285, 2019.
- [3] K. Hornik, M. Stinchcombe, and H. White, "Multilayer feedforward networks are universal approximators," *Neural Networks*, vol. 2, no. 5, pp. 359–366, 1989.
- [4] W. Ji and K. C. Chee, "Prediction of hourly solar radiation using a novel hybrid model of ARMA and TDNN," *Solar Energy*, vol. 85, no. 5, pp. 808–817, 2011.
- [5] H. Min, X. Jianhui, X. Shiguo, and Y. Fu-Liang, "Prediction of chaotic time series based on the recurrent predictor neural network," *IEEE Transactions on Signal Processing*, vol. 52, pp. 3409–3416, 2004.

- [6] X. Wu, Y. Wang, J. Mao, Z. Du, and C. Li, "Multi-step prediction of time series with random missing data," *Applied Mathematical Modelling*, vol. 38, no. 14, pp. 3512–3522, 2014.
- [7] X. Yi, J. Zhang, Z. Wang, T. Li, and Y. Zheng, "Deep distributed fusion network for air quality prediction," in *Proceedings of Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pp. 965–973, London, UK, 2018.
- [8] P.-H. Kuo and C.-J. Huang, "An electricity price forecasting model by hybrid structured deep neural networks," *Sustainability*, vol. 10, no. 4, p. 1280, 2018.
- [9] S. Mujeeb, N. Javaid, M. Ilahi, Z. Wadud, F. Ishmanov, and M. Afzal, "Deep long short-term memory: a new price and load forecasting scheme for big data in smart cities," *Sustainability*, vol. 11, no. 4, p. 987, 2019.
- [10] Z. Shen, Y. Zhang, J. Lu, J. Xu, and G. Xiao, "A novel time series forecasting model with deep learning," *Neurocomputing*, vol. 396, pp. 302–313, 2020.
- [11] M. Mudassir, S. Bennbaia, D. Unal, and M. Hammoudeh, "Time-series forecasting of bitcoin prices using high-dimensional features: a machine learning approach," *Neural Computing and Applications*, 2020.
- [12] D. Hussain, T. Hussain, A. A. Khan, S. A. A. Naqvi, and A. Jamil, "A deep learning approach for hydrological time-series prediction: a case study of gilgit river basin," *Earth Science Informatics*, vol. 13, no. 3, pp. 915–927, 2020.
- [13] V. K. R. Chimmula and L. Zhang, "Time series forecasting of covid-19 transmission in Canada using LSTM networks," *Chaos, Solitons & Fractals*, vol. 135, Article ID 109864, 2020.
- [14] I. E. Livieris, E. Pintelas, and P. Pintelas, "A CNN-LSTM model for gold price time-series forecasting," *Neural Computing and Applications*, 2020.
- [15] Z. Zhang and Y. Dong, "Temperature forecasting via convolutional recurrent neural networks based on time-series data," *Complexity*, vol. 2020, Article ID 3536572, 8 pages, 2020.
- [16] Y. Li, "Forecasting Chinese carbon emissions based on a novel time series prediction method," *Energy Science & Engineering*, vol. 8, no. 7, pp. 2274–2285, 2020.
- [17] R. Asadi and A. C. Regan, "A spatio-temporal decomposition based deep neural network for time series forecasting," *Applied Soft Computing*, vol. 87, Article ID 105963, 2020.
- [18] W. Peng, "DLI: a deep learning-based granger causality inference," *Complexity*, vol. 2020, Article ID 5960171, 8 pages, 2020.
- [19] B. Xin, F. Cao, W. Peng, and A. A. Elsadany, "A bertrand duopoly game with long-memory effects," *Complexity*, vol. 2020, Article ID 2924169, 7 pages, 2020.
- [20] B. Xin, W. Peng, and Y. Kwon, "A discrete fractional-order cournot duopoly game," *Physica A: Statistical Mechanics and its Applications*, vol. 558, Article ID 124993, 2020.
- [21] B. Xin and M. Sun, "A differential oligopoly game for optimal production planning and water savings," *European Journal of Operational Research*, vol. 269, no. 1, pp. 206–217, 2018.
- [22] B. Xin, W. Peng, Y. Kwon, and Y. Liu, "Modeling, discretization, and hyperchaos detection of conformable derivative approach to a financial system with market confidence and ethics risk," *Advances in Difference Equations*, vol. 2019, no. 1, p. 138, 2019.