A Convolutional Neural Network-based Prediction Mechanism for Sewage Treatment

Wenru Zeng, Zhiwei Guo, Jianhui Wang, Yu Shen and Xu Gao*
National Research Base of Intelligent Manufacturing Service, Chongqing Technology and Business University, Chongqing 400067, China

*Corresponding author

Abstract. Aiming at the extraordinary dependency on manual experience during the process of sewage treatment, it is required to explore intelligent prediction mechanism for sewage quality. Due to the fact that convolutional neural network (CNN) has some unique advantages for fast prediction, this paper proposes a convolutional neural network-based prediction mechanism for sewage treatment (CPST). It contains two mains parts respectively named learning process and training process. A group of experiments are managed to evaluate the efficiency of the proposed CPST, and the results demonstrate that CPST obtain the lower prediction error compared with baselines.

Keywords: Sewage treatment; Prediction methods; Convolutional neural network; Data-driven model.

1. Introduction
Sewage treatment controls the effluent water quality by adjusting dissolved oxygen in treatment process. Currently, sewage treatment is mainly manipulated with the aid of manual experience which is usually biased and results in waste of human resources. In fact, the quantity of dissolved oxygen and index of effluent water are highly related and can be associated using data-driven models. Deep learning has been proved to be a type of effective approaches to predict unknown things such as stock prices from the perspective of statistical learning[1,2]. Among, convolutional neural network (CNN) and recurrent neural network (RNN) are two most significant learning schemes[3,4]. In [5], Li et al. utilized the RNN structure-based LSTM model for this purpose. And Liu et al. presented an attention-based RNN model to achieve aquaculture-based dissolved oxygen prediction in [6]. However, compared with the RNN, the CNN is able to automatically extract features and possesses faster running speed[7,8]. Thus this article proposes a Convolutional neural network-based Prediction mechanism for Sewage Treatment (CPST). And major contributions of this paper are as follows:

- Recognition of feasibility of CNN in the process of sewage treatment;
- Proposal of CPST: a Convolutional Neural Network-based Prediction Mechanism for Sewage Treatment.

2. Overview

2.1. Problem Statement

Definition 1 (Inlet COD): Initial index of the chemical oxygen demand (COD) before entering into sewage treatment plant.

Definition 2 (Inlet NH$_4^+$-N): Initial index of the ammonia nitrogen (NH$_4^+$-N) before entering into sewage treatment plant.

Definition 3 (DO): The dissolved oxygen (DO) added during the treatment.
Definition 4 (Effluent COD): Final index of the chemical oxygen demand (COD) after completing sewage treatment processes.

Definition 5 (Effluent NH₄⁺-N): Final index of the ammonia nitrogen (NH₄⁺-N) after completing sewage treatment processes.

Taking a sewage plant as an example, the inlet and effluent water quality were monitored separately, and set up monitoring points in the aerobic tank of the A²/O process chain of the sewage treatment system. The goal is to predict the impact on the effluent water quality exerted by multiple A²/O process chains in the sewage treatment plant.

2.2. Framework

Figure 1 illustrates the framework of the proposed CPST, which consists of two parts: learning process and testing process.

**Figure 1.** Framework of the proposed CPST.

2.2.1. **Learning Process.** By comparing the correlation coefficients between features and water quality standards (COD and NH₄⁺-N in effluent), it is expected to set 0.1 as the threshold. Features with correlation coefficients greater than 0.1 in water quality standards are retained. The remaining features are then input into CNN model, core of the CPST. The CPST adoptively sets the convolutional layer, pooling layer, and fully connected layer to continuously learn the relationship between features, so as to better predict water quality.

Two benefits of using CNN in CPST are two-aspect: 1) The neurons on the same feature map have the same weight, so the network can learn in parallel[9]; 2) CNN is able to directly input multi-dimensional features into the network, escaping the intricacy of data restructurings during feature extraction and classification[10,11].

2.2.2. **Testing Process.** The training data is continuously trained through the above learning process until the accuracy reaches a higher level. The test data is then input into the trained model to test the stability and superiority of the model.
3. Methodology

3.1. Feature Extraction

This paper mainly studies the relationship among inlet water quality, dissolved oxygen concentration in six aerobic pools and effluent water quality, and further predicts the effluent water quality. The main indicators of this study are shown in Table 1.

### Table 1. The main indicators of this study.

| $x_{1}, x_{2}$       | COD and NH$_4^+$-N of inlet water quality |
|----------------------|------------------------------------------|
| $x_{3}, x_{4}$       | DO of the first series of unit A and unit B in aerobic pool |
| $x_{5}, x_{6}$       | DO of the second series of unit A and unit B in aerobic pool |
| $x_{7}, x_{8}$       | DO of the third series of unit A and unit B in aerobic pool |
| $y_{1}, y_{2}$       | COD and NH$_4^+$-N of Effluent water      |

#### 3.1.1. Data Pre-processing.

There is some amorosity with the initial data obtained from sewage treatment plants, such as the frequency inconsistency of daily monitoring, incidental damage to monitoring equipment and regular repair of equipment. Therefore, necessary preprocessing steps are conducted through following steps:

- **Controlling frequency inconsistency:** Due to frequency inconsistency of monitoring, we intercept the first 200 pieces of data from each day.

- **Missing value processing:** For data missing in a certain period of time on a certain day, the maximum and minimum values in the data within the first two hours are set as the range, and random numbers are added to ensure 200 pieces of data per day.

- **Data summary:** The processed data are summarized, and the data in each monitoring indicator is 365*200 (73000). Then statistical analysis and summary of this, as shown in Table 2.

### Table 2. Statistical analysis and summary of sewage treatment.

| Variables | Minimum | Maximum | Mean   | SD    | CV    |
|-----------|---------|---------|--------|-------|-------|
| $x_{1}$ (mg/L) | 9.339 | 1061.544 | 441.008 | 195.165 | 0.443 |
| $x_{2}$ (mg/L) | 0.156 | 110.467 | 27.283 | 9.889 | 0.362 |
| $x_{3}$ (mg/L) | 1.002 | 9.562 | 2.810 | 1.519 | 0.540 |
| $x_{4}$ (mg/L) | 1.000 | 9.287 | 3.293 | 1.861 | 0.565 |
| $x_{5}$ (mg/L) | 1.001 | 9.413 | 2.607 | 1.117 | 0.429 |
| $x_{6}$ (mg/L) | 1.000 | 9.088 | 2.691 | 1.109 | 0.412 |
| $x_{7}$ (mg/L) | 1.003 | 9.956 | 5.571 | 2.604 | 0.467 |
| $x_{8}$ (mg/L) | 1.241 | 9.973 | 6.152 | 3.118 | 0.507 |
| $y_{1}$ (mg/L) | 3.016 | 49.475 | 25.606 | 8.115 | 0.317 |
| $y_{2}$ (mg/L) | 0.156 | 28.607 | 2.215 | 1.558 | 0.703 |

SD: standard deviation.  
CV: coefficient of variation(SD/Mean).

#### 3.1.2. Correlation Analysis.

Table 3 shows the correlation coefficients ($R$) of the induced variables and water quality parameters. Correlation coefficients greater than 0.1 are selected as the threshold for selecting input variables. The results show that the most effective inputs affecting the effluent COD ($y_{1}$)
are $x_1$, $x_3$, $x_4$, $x_5$, $x_6$ and $x_8$. The most important dependent variables affecting the effluent NH$_4^+$-N ($y_2$) are $x_1$, $x_2$, $x_3$, $x_4$, $x_5$ and $x_8$.

Table 3. The correlation coefficient ($R$) between individual induced variables and $y_1$, $y_2$ respectively.

| Variables | $R$ value between dependent variable and $y_1$ | $R$ value between dependent variable and $y_2$ |
|-----------|---------------------------------|---------------------------------|
| $x_1$     | 0.284                           | 0.160                           |
| $x_2$     | 0.099                           | 0.142                           |
| $x_3$     | 0.417                           | 0.212                           |
| $x_4$     | 0.116                           | 0.139                           |
| $x_5$     | 0.286                           | 0.226                           |
| $x_6$     | 0.130                           | 0.098                           |
| $x_7$     | 0.058                           | 0.087                           |
| $x_8$     | 0.107                           | 0.118                           |

3.2. Convolutional Neural Network
LeCun with his coworkers recommended convolutional neural networks in 1995[12]. And CNN contains the following layers:

3.2.1. Convolutional Layer. The role of convolution layer is to perform convolution operations on the data. Assume that the input of layer $l-1$ is an $N \times N$ matrix, and use $F \times F$ filter. Then calculate the input of layer $l$ according to Eq.(1). Figure 2 shows the value of $v_{l,1}$ in the next layer by filtering the input data. Usually, the output after passing through the filter will pass through an activation function to the next layer. The activation function in CNN generally uses the Relu function (Eq.2).

$$v_{i,j}^{l} = \delta \left( \sum_{k=0}^{F-1} \sum_{m=0}^{F-1} w_{k,m} V_{i+k,j+m}^{l-1} \right)$$ (1)

$$f(x) = \max(0,x)$$ (2)

In Eq.1, $v_{i,j}^{l}$ is the value of the $i-th$ row and $j-th$ column of layer $l$, $w_{k,m}$ is the weight of the $k-th$ row and $m-th$ column in the filter, and $\delta$ is a non-linear activation function.
3.2.2. **Pooling Layer.** The pooling layer is also called the subsampling layer. It can not only compress the features to reduce the amount of calculation, but also prevent the overfitting problem to a certain extent.

3.2.3. **Fully Connected Layer.** The last layer of CNN is called fully connected layer, of which each node is connected to all the nodes in the previous layer, and is used to integrate the features extracted before. The relationship between two adjacent layers is defined by Eq.3:

\[ v^j_i = \delta \left( \sum_k w^{j+1}_{k,i} v^{j+1}_k \right) \]  

where \( v^j_i \) is the value of the \( i-th \) neuron in the \( j-th \) layer, \( \delta \) is the activation function, and \( w^{j+1}_{k,i} \) is the weight between connecting the \( k-th \) neuron in the \( j-1 \) layer and the \( i-th \) neuron in the \( j-th \) layer.

4. **Experiments**

4.1. **Dataset**

The data in this paper are collected from a sewage treatment factory based on A²/O process in Chongqing, China. The A²/O process, also known as A/A/O, is an anaerobic-hypoxic-aerobic biological treatment method. This is a common treatment process. It has good denitrification and dephosphorization effects through spatial changes, and can be used for secondary or tertiary sewage treatment and reclaimed water reuse[14,15]. The A²/O process process of the sewage treatment plant studied in this paper is shown as Figure 3.
4.2. Experimental Settings and Baselines
The independent data dimension is increased to 16, and then is reshaped into a two-dimensional 4*4 matrix. As for CNN network, two convolutional layers, one pooling layer, and one fully connected layer are set up. The activation function of the convolution layer and the pooling layer is the Relu function, the activation function of the fully connected layer is None. Besides, minimum loss function is set to the error function, with the learning rate setting as 0.01 and the operation number setting as 200.

We compare CNN with the following approaches:
- **MLR (Multilinear Regression):** It is a kind of regression analysis method using the least squares function to model the relationship between multiple independent variables [16].
- **MLP (Multilayer Perceptron):** It is a feedforward artificial neural network (ANN) model.
- **LSTM (Long Short-Term Memory):** It is a kind of RNN model, and is specifically designed to handle the long-term dependence problem of general RNN[17].

4.3. Metrics
In this paper, the following evaluation indicators are selected: mean absolute error (MAE), root mean square error (RMSE) and mean absolute percent error (MAPE). The formulas are as follows:

\[
MAE = \frac{1}{n} \sum_{i=1}^{n} |y_o^{(i)} - y_m^{(i)}|
\]  \hspace{1cm} (4)

\[
RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_o^{(i)} - y_m^{(i)})^2}
\]  \hspace{1cm} (5)
\[
MAPE = \frac{1}{n} \sum_{i=1}^{n} \left| \frac{y_o^{(i)} - y_m^{(i)}}{y_m^{(i)}} \right|
\]

Where \(y_o^{(i)}\) represents the observed price of \(y\) at the position \(i\), \(y_m^{(i)}\) represents the predicted value of \(y\) at the position \(i\), and \(n\) is the total number of data patterns used.

### 4.4. Results

In this part, the consequences of four diverse experiments are presented. Table 4, Figure 4, and Figure 5 show the MAE, RMSE, and MAPE of the four models, respectively. From the following chart, we can see that CPST's three indicators are better than MLR, MLP and LSTM. Therefore CPST can better handle complex multi-dimensional variable problems of sewage treatment prediction.

**Table 4. MAE comparison of four models.**

|                | MAE  | MLR  | MLP  | LSTM | CPST |
|----------------|------|------|------|------|------|
| Effluent COD   | 4.936| 3.986| 1.633| 0.485|      |
| Effluent NH₄⁻N | 1.994| 0.943| 0.636| 0.245|      |

**Figure 4. RMSE comparison of four models.**

**Figure 5. MAPE comparison of four models.**

### 5. Conclusion

This paper proposes the CPST framework to explore automatic control scheme of sewage treatment, and takes a sewage treatment plant in Chongqing as an example. By comparing with three reference models, the results show that the CPST perform better than baselines concerning three metrics: MAPE, RMSE and MAE.

**Acknowledgement**

This research was supported by the General Program of Chongqing Natural Science Foundation (cstc2019jcyj-msxmX0747) and the National Key Research & Development Program of China
(2016YFE0205600).

References

[1] K. Wang, K. Li, L. Zhou, Y. Hu, Z. Cheng, J. Liu, C. Chen, Multiple convolutional neural networks for multivariate time series prediction, Neurocomputing. 360 (2019) 107–119. https://doi.org/10.1016/j.neucom.2019.05.023.

[2] Y.R. Ding, Y.J. Cai, P.D. Sun, B. Chen, The use of combined neural networks and genetic algorithms for prediction of river water quality, J. Appl. Res. Technol. 12 (2014) 493–499. https://doi.org/10.1016/S1665-6423(14)71629-3.

[3] Z. Che, S. Purushotham, K. Cho, D. Sontag, Y. Liu, Recurrent Neural Networks for Multivariate Time Series with Missing Values, Sci. Rep. (2018). https://doi.org/10.1038/s41598-018-24271-9.

[4] C. Lea, R. Vidal, A. Reiter, G.D. Hager, Temporal convolutional networks: A unified approach to action segmentation, in: Lect. Notes Comput. Sci. (Including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), 2016. https://doi.org/10.1007/978-3-319-49409-8_7.

[5] Z. Li, F. Peng, B. Niu, G. Li, J. Wu, Z. Miao, Water Quality Prediction Model Combining Sparse Auto-encoder and LSTM Network, IFAC-PapersOnLine. 51 (2018) 831–836. https://doi.org/10.1016/j.ifacol.2018.08.091.

[6] Y. Liu, Q. Zhang, L. Song, Y. Chen, Attention-based recurrent neural networks for accurate short-term and long-term dissolved oxygen prediction, Comput. Electron. Agric. 165 (2019) 104964. https://doi.org/10.1016/j.compag.2019.104964.

[7] W. Liao, Y. Wang, Y. Yin, X. Zhang, P. Ma, Improved Sequence Generation Model for Multi-label Classification via CNN and Initialized Fully Connection, Neurocomputing. (2019). https://doi.org/10.1016/j.neucom.2019.11.074.

[8] P. Li, M. Abdel-Aty, J. Yuan, Real-time crash risk prediction on arterials based on LSTM-CNN, Accid. Anal. Prev. 135 (2020) 105371. https://doi.org/10.1016/j.aap.2019.105371.

[9] D. Konstantinidis, V. Argyriou, T. Stathaki, N. Grammalidis, A modular CNN-based building detector for remote sensing images, Comput. Networks. 168 (2020) 107034. https://doi.org/10.1016/j.comnet.2019.107034.

[10] E. Hoseinzade, S. Haratizadeh, CNNpred: CNN-based stock market prediction using a diverse set of variables, Expert Syst. Appl. 129 (2019) 273–285. https://doi.org/10.1016/j.eswa.2019.03.029.

[11] J. Su, C. Li, Q. Zeng, J. Yang, J. Zhang, A green closed-loop supply chain coordination mechanism based on third-party recycling, Sustain. 11 (2019) 1–14. https://doi.org/10.3390/su11195335.

[12] L. Name, F. Name, O. Training, P. Training, C. Darin, R.O. Training, M. Kimberly, G. Deepa, E. Board, E. Principal, I. Primary, F. Systems, E.B. Study, N. Co-investigator, Convolution Networks for Images, Speech, and Time-Series, Igarss 2014. (1998) 1–5. https://doi.org/10.1007/s13398-014-0173-7.2.

[13] Z.-R. Wang, J. Du, J.-M. Wang, Writer-aware CNN for parsimonious HMM-based offline handwritten Chinese text recognition, Pattern Recognit. 100 (2020) 107102. https://doi.org/10.1016/j.patcog.2019.107102.

[14] S. Jiafu, Y. Yu, Y. Tao, Measuring knowledge diffusion efficiency in R&D networks, Knowl. Manag. Res. Pract. 16 (2018) 208–219. https://doi.org/10.1080/14778238.2018.1435186.

[15] X. Zhang, J. Su, A combined fuzzy DEMATEL and TOPSIS approach for estimating participants in knowledge-intensive crowdsourcing, Comput. Ind. Eng. 137 (2019) 106085. https://doi.org/10.1016/j.cie.2019.106085.

[16] K. Abrougui, K. Gabsi, B. Mercatoris, C. Khemis, R. Amami, S. Chehaibi, Prediction of organic potato yield using tillage systems and soil properties by artificial neural network (ANN) and multiple linear regressions (MLR), Soil Tillage Res. 190 (2019) 202–208. https://doi.org/10.1016/J.STILL.2019.01.011.

[17] J. Su, Y. Yang, R. Duan, A CA-based heterogeneous model for knowledge dissemination inside knowledge-based organizations, J. Intell. Fuzzy Syst. 34 (2018) 2087–2097. https://doi.org/10.3233/JIFS-162116.