Multi-Task Temporal Convolutional Network for Predicting Water Quality Sensor Data

Yi-Fan Zhang[1] [0000−0002−9182−6267], Peter J. Thorburn[1] [0000−0002−6506−0456], and Peter Fitch[2] [0000−0002−9813−0588]

[1] Agriculture & Food, CSIRO, Brisbane, QLD, 4067, Australia
{Yi-Fan.Zhang,Peter.Thorburn}@csiro.au
[2] Land & Water, CSIRO, Canberra, ACT, 2601, Australia
Fitch.Peter@csiro.au

Abstract. Predicting the trend of water quality is essential in environmental management decision support systems. Despite various data-driven models in water quality prediction, most studies focus on predicting a single water quality variable. When multiple water quality variables need to be estimated, preparing several data-driven models may require unaffordable computing resources. Also, the changing patterns of several water quality variables can only be revealed by processing long term historical observations, which is not well supported by conventional data-driven models. In this paper, we propose a multi-task temporal convolution network (MTCN) for predicting multiple water quality variables. The temporal convolution offers one the capability to explore the temporal dependencies among a remarkably long historical period. Furthermore, instead of providing predictions for only one water quality variable, the MTCN is designed to predict multiple water quality variables simultaneously. Data collected from the Burnett River, Queensland is used to evaluate the MTCN. Compared to training a set of single-task TCNs for each variable separately, the proposed MTCN achieves the best RMSE scores in predicting both temperature and DO in the following 48 time steps by only requires 53% of the total training time. Therefore, the MTCN is an encouraging approach for water quality management by processing a large amount of sensor data.

Keywords: Prediction Model · Multi-Task Learning · Water Quality.

1 Introduction

Water quality is one of the major issues today because of its effects on human health and aquatic ecosystems. The water quality deterioration can be attributed to urbanisation, population growth, excessive water consumption, industrial wastewater discharge, and agricultural activities in the catchments [5]. An understanding of water quality dynamics is critical to the intelligent decision making in regards to ecological conservation [12].

Capturing long-term dependencies in time series data remains a fundamental challenge [17, 16]. Despite advances in building models based on recurrent neural networks (RNNs), those models are still difficult to scale to very long data sequences. In the study proposed by Wang et al. [14], the maximum number of historical time steps used in their dissolved oxygen predictive model is five. Moon et al. [11] proposed an RNN-based model in forecasting electrical conductivity. Their model achieved the best performance when processing the inputs from 24 previous timesteps. Inputs with a small number of timesteps limits these RNN-based predictive models in identifying the long-range changing patterns, which is critical in numerous water quality variables.

Temporal Convolutional Networks (TCNs) overcome the previous shortcomings by capturing long-range patterns using a hierarchy of temporal convolutional filters [7]. Instead of using recurrent structure to maintain temporal dependencies, the TCN applies various sizes of convolutional filters to obtain the temporal dependencies at different time scale. Also, the dilated convolutions [13] increase the receptive field significantly so long historical data can be utilised.

Furthermore, most water quality researchers build predictive models for single water quality variable. For example, Alizadeh et al. [2] applied 30 different artificial neural network (ANN) models to predict daily values of salinity, temperature and DO separately. In the study proposed by Kim and Seo [6], an ANN ensemble model was developed to forecast the water quality variables such as pH, DO, turbidity, total nitrogen and phosphorus. The ANN ensemble model included 150 individual ANN models, with each of them needing to be trained and evaluated. In these studies, though all the models indeed deal with the same datasets, they cannot obtain benefits from each other's learning process.

Multi-task learning is an essential machine learning paradigm which aims at improving the generalisation performance of a task by using other related tasks [8]. It is prevalent in various applications ranging from computer vision [10] to speech recognition [4]. In the context of water quality prediction, each water quality variable interacts with and influences other variables in the same ecosystem. The temporal patterns of one water quality variable can, therefore, be precious in guiding us predicting other water constituents' values.

In this paper, we propose a multi-task temporal convolution network (MTCN) for predicting multiple water quality variables. The key contributions include:

- We develop a multi-variable predictive model to forecast various water quality constituents simultaneously. Applying a unified model in predicting multiple variables enables the knowledge sharing between multiple learning processes, and also reduces the necessities of computing resources significantly.
- We applied the temporal convolution network (TCN) to learn the long-term temporal dependencies for water quality data. Comparing to the RNN-based models, the TCN exhibits longer effective history data than the recurrent counterparts. The experiments demonstrate that the MTCN can obtain superior performance compared to equivalent separately trained models.

2 Proposed Multi-Task Temporal Convolution Network

In this section, we propose a water quality multi-variable predictive model for forecasting various water quality constituents simultaneously. A TCN-based predictive model is built by following the multi-task learning paradigm. The model is designed to learn the temporal dependencies among various water quality monitoring properties within a long period of time. Each predictive task can benefit from the shared hidden representations. Moreover, task-specific layers are assigned to forecast the corresponded water quality variable concurrently.

We implement a temporal convolutional network similar to the one proposed by Bai et al. [3]. The TCN includes a stack of causal convolutional layers. Causal convolution is used to make sure the model will not capture information from the future time index to help the prediction task. In addition, the dilated convolutions and the residual connections are integrated into the TCN to enhance the utilization of long historical observations without the vastly deep structure.

2.1 Dilated Convolution

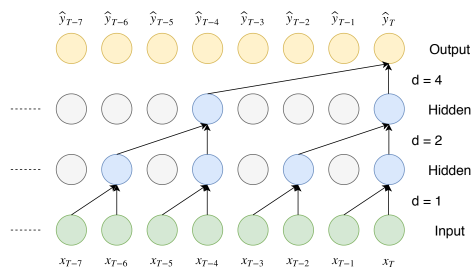


Fig. 1. The TCN with dilated convolutions. The dilated causal convolution is defined with dilation factors $d = [1, 2, 4]$, and filter size $k = 2$. In this case, the TCN is able to cover 8 numbers of historical observations.

Figure 1 illustrates the way of applying dilated convolutions to increase the size of the receptive field. The dilated convolution operator can apply the same filter at different time scales using different dilation factors. 1D dilated convolution is defined as:

$$g[i] = \sum_{l=1}^{L} f[i + d \cdot l]h[l], \tag{1}$$

where $f[i]$ and $g[i]$ are the input and output time series, $h[l]$ denotes the filter of length L and d corresponds to the dilation rate.

2.2 Residual Unit

A residual unit defined in Bai et al.'s study [3] is implemented to improve the TCN's stability. The residual block (Figure 2) includes two dilated causal convolutional layers. The weight normalization is applied to the convolutional filters and a spatial dropout was added after each dilated convolution for regularization. In addition, the input of the residual unit is added to the output through an additional 1×1 convolution.

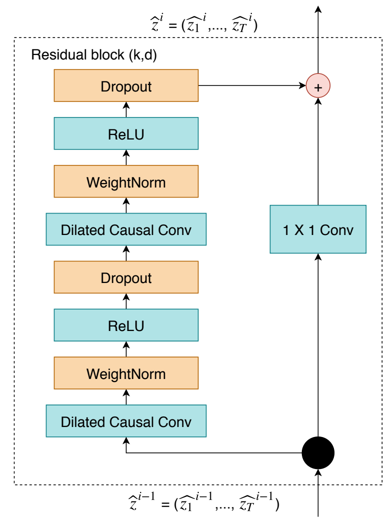


Fig. 2. Residual unit in the TCN. With the help of the skip connection within the residual unit, one can propagate larger gradients through the neural network.

2.3 Multi-Task Temporal Convolution Network

In this subsection, we developed our multi-task temporal convolution network (MTCN) based on the TCN and the multi-task learning paradigm.

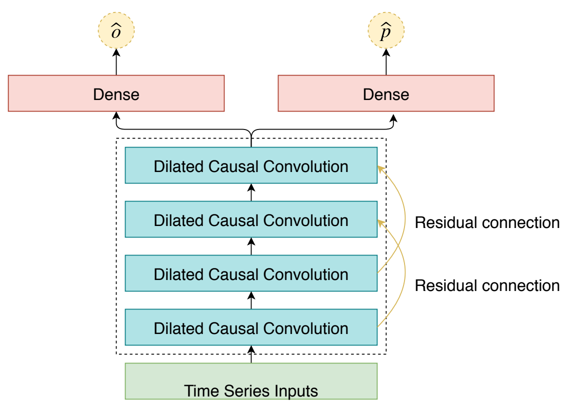


Fig. 3. The proposed MTCN. $\hat{o}$ and $\hat{p}$ represent the predictions of different variables.

The proposed MTCN is illustrated in Figure 3. By adjusting the dilation factors and filter size, the MTCN can cover a wide range of time series data by applying a hierarchy of filters with various sizes. In addition, the residual connections help to maintain the stability of the deep neural network by enhancing the information flow through the initial layer to the last layer in the deep neural network. The task-specific dense layers with the linear activation function are added on top of the shared convolutional layers.

3 Evaluation

In this section, we evaluate the effectiveness of the MTCN by using the water quality data collected by a water quality monitoring program in Australia.

3.1 Water Quality Sensor Data

The Burnett River is located on the southern Queensland coast and flows into the coral sea of the South Pacific Ocean. Cultivation of sugarcane and small crop are important lands uses in this region. The YSI model 6-Series Sonde is deployed and monitoring the water quality variables [1]. Temperature, electric conductivity (EC), pH, dissolved oxygen (DO), turbidity and chlorophyll-a (Chl-a) are recorded with half an hour time interval (Table 1).

Table 1. Water quality data during 1/3/2014 and 31/3/2018.

Variables	Unit	Min	Max	Mean	Std Dev
Temperature	$^{\circ}$C	13.60	32.71	24.64	3.96
Electrical conductivity	μS / cm	2	50720	35931.09	14428.10
pH		6.62	8.63	7.85	0.63
Dissolved oxygen	mg L^{-1}	2.06	13.90	6.64	0.98
Turbidity	NTU	0.1	1850	19.85	87.18
Chlorophyll-a	μg L^{-1}	0.1	345.60	10.07	31.89

We choose the sensor data from 1/3/2014-31/3/2017 as training data and sensor data from 1/4/2017-31/3/2018 as testing data. During the training, 10% samples are selected as validation data. Considering the missing and abnormal measurements are inevitable in the monitoring network, we cleaned and normalized the chosen datasets first before feeding into the neural network models.

Beside this, studies [9, 15] confirm that the concentration of DO in surface water is controlled by temperature and has both a seasonal and a daily cycle. We designed an MTCN to predict DO and temperature simultaneously. Two comparative TCNs were also designed to forecast the DO concentration and temperature separately.

3.2 Experimental Settings

To measure the performance of the predictive model, we used the mean absolute error (MAE) and the root mean square error (RMSE). Also, some of the optimised key hyperparameters are listed in Table 2.

Table 2. Key hyperparameters of the MTCN.

Hyperparameters	Value
No. of Dense Layers (per task)	2
No. of Units in Dense Layers (per task)	$[64, 48]$
Dilated Factors	$[1, 2, 4, 8, 16, 32, 64]$
Kernel Size	3
Dropout Rate	0.6

Based on the dilated factors and filter's kernel size (Table 2), the MTCN can cover 192 historical observations for predicting both the temperature and dissolved oxygen values in the future 48 time index. According to this experimental design, the MTCN is able to forecast the changing of the temperature and dissolved oxygen in the following 24 hours.

3.3 Experimental Results and Discussion

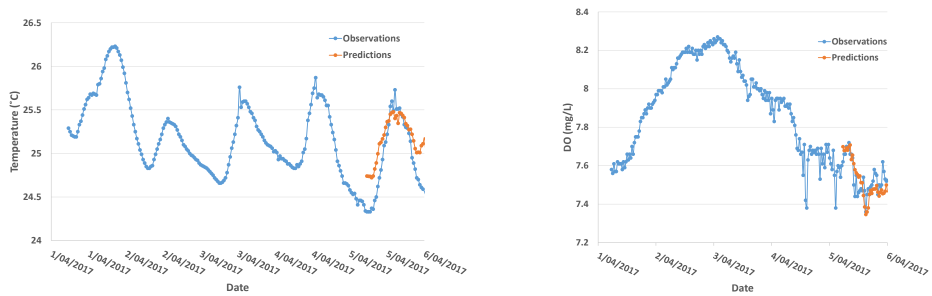
We also compare the MTCN with the single-task TCN. The single-task TCN shares the same hyperparameter setting with MTCN, while it does not have the task-specific dense layer and multiple outputs. Hence, multiple TCNs have to be trained to meet the requirements of multi-variables prediction.

Table 3. Performance Measurement.

Model	Metrics	Prediction Accuracy		Training Time
		Temperature	DO	
MTCN	RMSE	**0.59**	**0.49**	**9H:58M**
	MAE	**0.37**	0.27	
TCN	RMSE	0.60	0.49	18H:49M
	MAE	0.38	**0.26**	

Table 3 illustrates model performance for both MTCN and TCN. Benefiting from the temporal convolutional architecture, dilated convolution and the residual unit, both MTCN and TCN achieve remarkable predictive accuracy for both DO and temperature. As shown in Figure 4a, the MTCN captures the trend of DO in the following 24 hours, and also gives the proper estimation when the concentration of DO drops significantly. Similarly, the temperature predictions generated by the MTCN follow the expected daily temperature variation in Figure 4b. Furthermore, the MTCN gains the best performance of both RMSE and MAE in predicting the change of temperature. Similarly, the MTCN also achieves the best RMSE scores in predicting the trend of DO.

In addition, the MTCN implemented in this experiment includes 1,304,096 trainable parameters, while the TCN with a single prediction task only has



(a) Predicting the trend of DO in 5/4/2017.

(b) Predicting the trend of temperature in 5/4/2017.

Fig. 4. Predicting the trend of DO and temperature by using the MTCN. 48 predictions are generated every half an hour based on historical inputs data.

710,064 parameters to train. To make a fair comparison, the training process was stopped after the 900th epoch for all the models listed in Table 3. Larger hidden parameters in the MTCN indicates that it requires more training epochs to converge, while the total training time is still much less than training separate TCNs for individual tasks. Therefore, the MTCN offers an efficient way in building predictive model for a number of water quality variables.

4 Conclusion

The development of reliable water quality predictions is critical to improve the management of aquatic ecosystems. This paper proposed a multi-task temporal convolutional network for predicting multiple water quality variables simultaneously. Experimental results were presented to demonstrate that the proposed model can achieve promising predictive accuracy for long term water quality prediction while requiring a significantly reduced training time.

Acknowledgement

This work was conducted within the CSIRO Digiscape Future Science Platform.

References

1. Ambient estuarine water quality monitoring data. `https://data.qld.gov.au/dataset`, accessed: 2017-11-20
2. Alizadeh, M.J., Kavianpour, M.R.: Development of wavelet-ANN models to predict water quality parameters in hilo bay, pacific ocean. Marine Pollution Bulletin **98**(1-2), 171–178 (2015)

3. Bai, S., Kolter, J.Z., Koltun, V.: An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. arXiv preprint arXiv:1803.01271 (2018)
4. Huang, J.T., Li, J., Yu, D., Deng, L., Gong, Y.: Cross-language knowledge transfer using multilingual deep neural network with shared hidden layers. In: Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on. pp. 7304–7308. IEEE (2013)
5. Ji, X., Shang, X., Dahlgren, R.A., Zhang, M.: Prediction of dissolved oxygen concentration in hypoxic river systems using support vector machine: a case study of wen-rui tang river, china. Environmental Science and Pollution Research **24**(19), 16062–16076 (2017)
6. Kim, S.E., Seo, I.W.: Artificial neural network ensemble modeling with conjunctive data clustering for water quality prediction in rivers. Journal of Hydro-environment Research **9**(3), 325–339 (2015)
7. Lea, C., Flynn, M.D., Vidal, R., Reiter, A., Hager, G.D.: Temporal convolutional networks for action segmentation and detection. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 1003–1012. IEEE (2017)
8. Luong, M.T., Le, Q.V., Sutskever, I., Vinyals, O., Kaiser, L.: Multi-task sequence to sequence learning. arXiv preprint arXiv:1511.06114 (2015)
9. Manasrah, R., Raheed, M., Badran, M.I.: Relationships between water temperature, nutrients and dissolved oxygen in the northern gulf of aqaba, red sea. Oceanologia **48**(2) (2006)
10. Misra, I., Shrivastava, A., Gupta, A., Hebert, M.: Cross-stitch networks for multi-task learning. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 3994–4003 (2016)
11. Moon, T., Ahn, T.I., Son, J.E.: Forecasting root-zone electrical conductivity of nutrient solutions in closed-loop soilless cultures via a recurrent neural network using environmental and cultivation information. Frontiers in plant science **9** (2018)
12. Thorburn, P., Wilkinson, S.: Conceptual frameworks for estimating the water quality benefits of improved agricultural management practices in large catchments. Agriculture, Ecosystems & Environment **180**, 192 – 209 (2013)
13. Van Den Oord, A., Dieleman, S., Zen, H., Simonyan, K., Vinyals, O., Graves, A., Kalchbrenner, N., Senior, A., Kavukcuoglu, K.: Wavenet: A generative model for raw audio. CoRR abs/1609.03499 (2016)
14. Wang, Y., Zhou, J., Chen, K., Wang, Y., Liu, L.: Water quality prediction method based on lstm neural network. In: 12th International Conference on Intelligent Systems and Knowledge Engineering (ISKE). pp. 1–5. IEEE (2017)
15. Zhang, Y., Fitch, P., Vilas, M.P., Thorburn, P.J.: Applying multi-layer artificial neural network and mutual information to the prediction of trends in dissolved oxygen. Frontiers in Environmental Science **7**, 46 (2019)
16. Zhang, Y., Fitch, P., Vilas, M.P., Thorburn, P.J.: Predicting the trend of dissolved oxygen based on kpca-rnn model. Water (2019), submitted
17. Zhang, Y., Thorburn, P.J., Wei, X., Fitch, P.: SSIM -a deep learning approach for recovering missing time series sensor data. IEEE Internet of Things Journal **6**(4), 6618–6628 (2019)