Forecasting the Color and Ammonia Concentration in the Reclaimed Water using Deep Learning

by

Ting Hsi LEE

A Thesis Submitted to
The Hong Kong University of Science and Technology
in Partial Fulfillment of the Requirements for
the Degree of Master of Philosophy
in the Department of Civil and Environmental Engineering

July 2022, Hong Kong

Authorization

I hereby declare that I am the sole author of the thesis.

I authorize the Hong Kong University of Science and Technology to lend this thesis to other institutions or individuals for the purpose of scholarly research.

I further authorize the Hong Kong University of Science and Technology to reproduce the thesis by photocopying or by other means, in total or in part, at the request of other institutions or individuals for the purpose of scholarly research.

Ting Hsi LEE

July 2022

Forecasting the Color and Ammonia Concentration in the Reclaimed Water using Deep Learning

by

Ting Hsi LEE

This is to certify that I have examined the above MPhil thesis
and have found that it is complete and satisfactory in all respects,
and that any and all revisions required by
the thesis examination committee have been made.

Prof. Chii SHANG, Thesis Supervisor

Prof. Meimei Han, Head of Department

Department of Civil and Environmental Engineering
July 2022

Acknowledgments

First of all, I am truly grateful for being one of the first PhD students supervised by Prof. Li. He was full of passion and patience when helping me build the know-how for this degree. It has been a great pleasure for me to be part of this team and grow together with the lab during the last four years. Furthermore, I would like to thank all of the members of the thesis examination committee for their careful examination of my thesis.

Finally, I would not stand at this current point without their endless love and unconditional support for all these years.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Forecasting the Color and Ammonia Concentration in the Reclaimed Water using Deep Learning

by Ting Hsi LEE

Department of Civil and Environmental Engineering

The Hong Kong University of Science and Technology

Abstract

Water scarcity is a global challenge. One of the promising ways to mitigate the water resource crisis is via wastewater reclamation. Chlorine is commonly used for reclaimed water disinfection and requires precise dosing to satisfy endorsed quality standards. Ammoniacal nitrogen (NH_3N) and colour exist in the reclaimed water at concentrations between 0.23 – 5.44 mg N/L and 80 – 150 Hazen units, respectively, and can affect the chlorine demand. Forecasting the reclaimed water quality enables a feedback control system over the disinfection process by predicting the exact chlorine dose required which secures sufficient time to respond to sudden surges in color and ammonia levels. This study developed time-variant models based on machine learning to predict the NH_3N concentration and colour three hours into the future in the reclaimed water. The NH_3N data was collected by an online analyzer, and colour data was collected by a customized auto-sampling spectrophotometer, both are installed in the reclaimed water treatment plant in Hong Kong. Long Short-Term Memory (LSTM) was found to be the most effective architecture for training NH_3N and colour forecasting models. In the training processes, we applied data pre-processing methods and feature engineering, a technique to select or create relevant variables in raw data to enhance predictive model performance. From feature engineering, we discovered that the daily fluctuation in NH_3N and colour has correlations with the urban water consumption patterns. This finding further enhanced the NH_3N and colour forecasting model performance by 4.9% and 5.4% compared to baseline models. This research work offers novel methods and feature engineering pro-

cesses for NH_3N concentration and colour forecasting in reclaimed water for treatment optimization.

CHAPTER 1

INTRODUCTION

1.1 Background

AI technologies have been successfully applied to different DWT processes, such as the prediction of the coagulant dosage, discrimination of the DBP formation potential, advanced control of membrane fouling, membrane preparation and optimization, and water quality prediction. [2]

Forecasting models play an important roles in water quality control in drinking water treatment plants (DTPs) and wastewater treatment plants (WWTPs). The need of using forecasting models are becuase the unpredictable nature of water quality, and the treatment operations are subjected to the change of water quality to prodcue effluent complied the government regulation [1].

Forecasting models can also be called time series model becuase the data is consisted of the values and the time (need to be further revised). For the well-know time series models are for example, RNN, ... These are used to replace the theory-based models, for example Activated Sludege Model (ASM). The difference between these two models are, machine learning based models require to learn from historic data, while the thoery-based models only need to enter the basic operational parameters (e.g., influent flow, tempearture, and pH, etc).

Despite the promising usage and performance of machine learning models, the collection of the data became the most difficult tasks. Many small scale or old treatment plants do not have the capital or the available environment for the set-ups of the online sensors to collect data. Although these are the major issues, it's still possible to train a forecasting model with one input, which is also called a self-prediction model. Although the accuracy or stability compared to multi-input models, the forecasted results can be used at some cases. To increase the model performance, there are several ways. Paper included weather data, or perform data-preprocessing methods to improve the model performance.

These solutions (data preprocessing, feature engineering) are not well discussed in this field, also the potential of using univariate models are under estimated.

1.2 Objectives

The specific objectives of this thesis work are:
(1) To build baseline univariate forecasting models using machine learning and deep learning models.
(2) To develop data preprocessing methods for enhancing model forecasting performance.
(3) To extract features and hidden relations of water parameters in MBR effluent by analyzing the wastewater collected upstream of the WWTPs.
(4) To develop methods for improving performance of forecasting models using the hidden features and relations of the water parameters.

1.3 Organization of the thesis

CHAPTER 2

LITERATURE REVIEW

2.1 The application of machine learning techniques for water quality control

2.1.1 Automated control system for water quality control

A proportional–integral–derivative controller (PID controller or three-term controller) is a control loop mechanism employing feedback that is widely used in industrial control systems and a variety of other applications requiring continuously modulated control. A PID controller continuously calculates an error value e(t) as the difference between a desired setpoint (SP) and a measured process variable (PV) and applies a correction based on proportional, integral, and derivative terms (denoted P, I, and D respectively), hence the name.

In practical terms, PID automatically applies an accurate and responsive correction to a control function. An everyday example is the cruise control on a car, where ascending a hill would lower speed if constant engine power were applied. The controller's PID algorithm restores the measured speed to the desired speed with minimal delay and overshoot by increasing the power output of the engine in a controlled manner.

In general terms PLCs are probably one of most widely used pieces of control and automation technology. The clue really comes from the name PLC, or "programmable logic controller". It is the fact that they are programmable that makes them so versatile in their application. PLCs contain a processor, memory to hold their programming and other data and input and output modules. They are usually programmed via a PC and there are a number of different industry standard (IEC 61131-3) languages that may be used.

A PID Controller is different to a PLC. It still requires inputs and outputs to receive information from the process and send signals back to control it but it contains specialist algorithms designed to control a process with one or multiple control loops. The term 'PID' relates to "Proportional Integral Derivative" control.

So that the machine learning nowadays focus on how to develop a better mathematical algorithms to replace the outdated algorithms in PID system.

To ensure the concentration of free residual chlorine falls in the recommended range, chlorination process has been made automated to correct the chlorine dosage responding to the variability in water quality in many DWTPs.

To operate a automatic control system, Programmable Logic Controller (PLC) is selected by industrial processes that involve in any sorts of control systems. In DWTPs, PLC system in chlorination process makes automatic decisions for the administration of chlorine dose required to achieve desirable concentration of FRC. One of the earliest mathematical modeling used with PLC to control chlorine dosage is proposed by Dieu, Garrett Jr., Ahmad, and Young (1995), which is a single variable Proportional-Integral Derivative (PID) chlorine dosage control system. PID system is built with online analyzer of FRC, PLC, and chlorine dosers to formulate a robust control loop for dosing chlorine. However, PID system is merely a temporarily solution to the precise control of chlorination.

Since the WTP postchlorination dosage has several variables that directly impact the FRC rate in the treated water, the use of the traditional PID controller is restricted, as reported by Escobar and Trierweiler (2013).

Machine learning is a subset of artificial intelligence, and deep learning is a subset of machine learning. In artificial intelligence can be sued to solve four types of problems: classification, regression, dimensionality reduction and clustering.

2.1.2 Water quality control in drinking water treatment plants

2.1.2.1 Disinfection process

Disinfection is the last step of water treatment processes in drinking water treatment plants (DWTPs) to generate safe potable water. In this step, one or more chemical disinfectants like chlorine, chloramine, or chlorine dioxide are added into the water to inactivate any remaining pathogenic microorganisms. The residual disinfectant concentration in disinfected water must contains low levels of the chemical disinfectant to stop nuisance growths in the water distribution pipes, storage facilites and conduits. Nowadays, the widely used disinfectant in disinfection process is chlorine as gas or hypochlorite (i.e., in form of liquid solution), and the treatment process is known as "chlorination".

According to World Health Organization's Guidelines for Drinking-water Quality (WHO Guidelines), the maximum allowable value for free chlorine residual in drinking water is 5 mg/L, and the minimum recommended valus is 0.2 mg/L.

Current analysis proposes a multivariable control for post-chlorination dosage system in a WTP using artificial neural networks applied to the disinfection process to reduce free residual chlorine variations of treated water in the water tank and, consequently, in the main water distribution [3].

Despite the benefit brought by dosing chlorine to the water, negative impacts also come along. In the real world case, the influent water quality and the efficiency of the drinking water treatment processes are not always stable, and the invariability of the treated water quality becomes an big issue for disinfection. For instance, chlorine dose can be excessive dosed when the treated water contains less pollutants (e.g., non-organic matters and ammonia nitrogen). Although the quality of disinfected water fulfill the regulation standard, it increases the costs and can potentially generate undesired disinfection by-products (e.g., trihalomethanes, which are carcinogenic to human) due to the chemcial reaction between pollutants and overly dosed chlorine. On the flip side, insufficient dosing of chlorine cuases the concentration of residual chlorine lower than the legal regulation. To prevent both senarios occur, a water quality control strategy is required to produce drinking water with satisfactory quality.

Up untill present, there are several ways to perform disinfected water quality. In the earliest time, feed-back.... PI... feed-foward...

2.1.2.2 Membrane fouling

Madfs

2.1.2.3 Analysis of precursors of DBPs

2.1.2.4 Disinfection

2.1.2.5 Prediction of the source water contaminants

2.1.2.6 Coagulation

Traditional modelling methods mainly use numerical simulations or physical formulas to model target prediction objects from a microscopic perspective . For example, the advantage of particle coagulation dynamics simulation is that it can ex plain the behaviour evolution mechanism of particles in the water treatment process in a very specific way because it is usually based on the collision mechanism with physical meaning and mathematical description

2.1.3 Water quality control in wastewater treatment plants

2.1.4 Water quality control in reclaimed water system

In this study the new control objectives for the reclaimed water system in Shek Wu Hui Effluent Polish Plant have been established: to monitor color and ammonia concentration in the MBR effluent and at the same time provide a predictive model to assist the disinfection control strategy for disinfecting the MBR effluent to meet the endorsed reclaimed water standard.

2.2 Recent advances in time series models for water quality forecasting

2.2.1 Machine learning models

2.2.2 Deep learning models

2.2.3 Comparison of the artificial intelligence model and traditional model in drinking water treatment

2.2.3.1 Traditional modeling methods

In traditional modeling methods, numerical simulations or physical formulas to model target prediction objects. ame training set to process batches and feed batches for ultrafiltration. The interpretation of this model is more accessible than simply using ANN because it is based on physical mechanisms. A semi-physical model can be defined as an aid to a mechanism because it provides an efficient way to determine specific parameters. Nevertheless, its further applications are limited by the assumptions established by the model.

2.3 Different techniques for enchancing the performance of forecasting models

2.3.1 Data preprocessing

2.3.2 Feature engineering

REFERENCES

[1] J.C. Chen, N.B. Chang, and W.K. Shieh. Assessing wastewater reclamation potential by neural network model. *Engineering Applications of Artificial Intelligence*, 16(2):149–157, March 2003.

[2] Lei Li, Shuming Rong, Rui Wang, and Shuili Yu. Recent advances in artificial intelligence and machine learning for nonlinear relationship analysis and process control in drinking water treatment: A review. *Chemical Engineering Journal*, 405:126673, February 2021.

[3] André Felipe Librantz, Fábio Cosme Rodrigues dos Santos, and Cleber Gustavo Dias. <b>Artificial neural networks to control chlorine dosing in a water treatment plant. *Acta Scientiarum. Technology*, 40(1):37275, September 2018.