Available online at www.sciencedirect.com

SciVerse ScienceDirect

Procedia Engineering 15 (2011) 5558 – 5562

Procedia Engineering

www.elsevier.com/locate/procedia

Advanced in Control Engineering and Information Science

KPCA and ELM ensemble modeling of wastewater effluent quality indices

L.J. Zhao[a, b]*, D. C. Yuan[a], T. Y. Chai[b], J. Tang[b]

[a] *Province Key Laboratory of Process Control Technology for Chemical Industry, Shenyang University of Chemical Technology, ,Shenyang, 11004, China*

[b]*State Key Laboratory of Synthetical Automation for Process Industries, Northeastern University,Shenyang, 110004, China*

Abstract

Reliable measurements of effluent quality are important for different operational tasks such as process monitoring, online simulation, and advanced control in the wastewater treatment process (WWTP). A kernel principal component analysis (KPCA) and extreme learning machine (ELM) based ensemble soft sensing model for effluent quality prediction was proposed. KPCA was used to extract nonlinear feature of input space to overcome high dimension and colinearity. ELM algorithm is inserted into the ensemble frame as a component model since ELM runs much faster and provides better generalization performance than the other popular learning algorithm. The average output of all the ELM components in the ensemble is the final estimation of the effluent quality index. Simulations results using industrial process data show that the reliability and accuracy based KPCA and ELM ensemble soft sensing outperform the ELM, ELM ensemble model.

Keywords: Extreme learning machine, ensemble modeling, KPCA, wastewater treatment;

1. INTRODUCTION

Due to lack of reliable on-line sensors to measure effluent quality parameters, it is difficult to implement monitoring, control and operational optimization in the wastewater treatment plants. However, existing on-line hardware sensors is not sufficient in accuracy and reliability due to big investments, poor reliability, and difficult maintenance [1]. Data-driven soft sensors have gained increased popularity with

* Corresponding author. Tel.: +86-024-89383907
E-mail address: zlj_lunlun@163.com.

1877-7058 © 2011 Published by Elsevier Ltd.
doi:10.1016/j.proeng.2011.08.1031

the availability of recorded and stored data. The most popular modeling techniques, such as multivariate statistical method, Artificial Neural Networks, and Support Vector Machines (SVMs), are successfully applied to monitor and model data-driven soft sensor[2-3].

Most of the traditional data-driven modeling methods are based on a single model, which is sometimes difficult to achieve a satisfactory performance for the complex processes with strong nonlinearity, time-variant and highly uncertainty. KPCA is a well-known method for nonlinear feature extraction[4]. Ensemble methods have received special attentions because it can improve the accuracy of predictor and achieve better stability through training a set of models and then combining them for final predictions [5]. The speed of neural networks is in general far slower than required and it has been a major bottleneck in their applications for past decades. Recently, Huang et al. proposed a new learning algorithm called Extreme Learning Machine (ELM) which randomly all the hidden nodes parameters of generalized Single-hidden Layer Feedforward Networks (SLFNs) and analytically determines the output weights of SLFNs[6]. In this study, a KPCA and ELM based ensemble soft sensing was proposed for predicting the effluent quality in the municipal treatment plants.

2. KPCA-ELM Ensemble Modeling

Influent quality and quantity, operating conditions and external environment change affect effluent quality in the wastewater treatment process. Therefore, influent quality and flow rate, returned sludge flow, aeration flow rate flow bioreactor, oxidation-reduction potential (ORP) and dissolved oxygen (DO) in the anoxic and aerobic tank, mixed liquor suspended solid (MLSS), pH, sludge volume (SV) and sludge volume index (SVI) at the bioreactor are chosen as inputs of soft-sensing model. The outputs of the soft sensor are effluent quality indices, such as biochemical oxygen demand (BOD_5) and ammonia nitrogen (NH). The structure of the ensemble soft sensing based on KPCA and ELM is shown in Fig. 1.

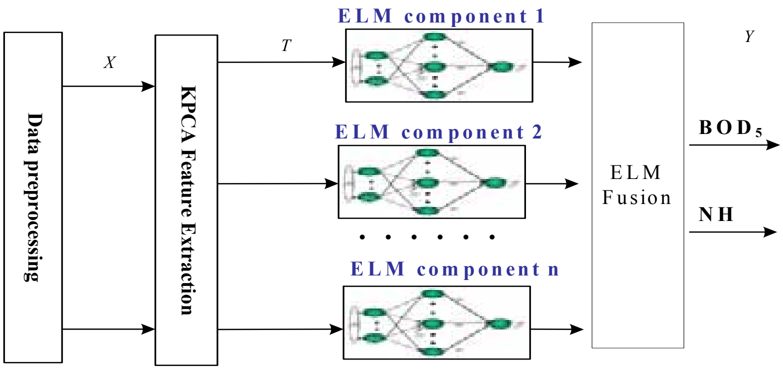

Fig. 1 Strategy of the ensemble soft sensing based on KPCA and ELM

It consists of the following parts: the data preprocessing, KPCA feature extraction, ELM component and ELM fusion. The prediction capacity of data-driven model strongly depends on the quality of the training data. So, gross errors and missing data contained in the history data are first processed based on a robust EMPCA method. Then, KPCA is used to eliminate the colinearity and noise in the data. ELM algorithm was inserted into the ensemble frame as a component model since ELM runs much faster and provides better generalization performance than other popular learning algorithm, which may overcome variations in different trials of simulations for a single ELM model. Finally, the average value of outputs of the ELM ensemble was the final measurements.