Online Sequential Prediction of Minority Class of Suspended Particulate Matters by Meta-Cognitive OS-ELM

by

Chiu Chi Chong

M-B0-5460-0

Master of Science in Software Engineering

2013

Faculty of Science and Technology
University of Macau

University of Macau

Abstract

ONLINE SEQUENTIAL PREDICTION OF MINORITY CLASS OF SUSPENDED PARTICULATE MATTERS BY META-COGNITIVE OS-ELM

by Chiu Chi Chong

Thesis Supervisors:
Dr. Vong Chi Man & Dr. Ip Weng Fai

Suspended particulate matters (PM_{10}) is considered as a harmful air pollutant. Many models attempt to predict numerical levels of PM_{10} but a simple, clearly defined classification of PM_{10} levels is more readily comprehensible to the general public rather than a numerical value. However, PM_{10} prediction model (e.g., support vector machine (SVM)) often suffers from data imbalance problem in the training dataset that results in failure to forecast the minority class of severe cases. In this thesis, a warning system using extreme learning machine (ELM), compared with SVM, was constructed to forecast the class of PM_{10} level: *Good*, *Moderate*, and *Severe*. An imbalance strategy called prior duplication was also applied to improve the forecast of minority class. The experimental comparisons between ELM and SVM demonstrate that ELM produces superior accuracy relative to SVM in forecasting minority class (*Severe*) of PM_{10} level with or without the imbalance strategy. Furthermore, the experimental results show that the required training time and model size in the ELM model are much shorter and smaller than those of SVM respectively, leading to a more efficient and practical implementation of prediction model for large dataset. However, ELM is a batch learning algorithm and many time-series problems such as air pollution index forecast require online sequential learning rather than batch learning. Therefore, this thesis proposes a new method called *meta-cognitive online sequential extreme learning machine* (MCOS-ELM) that aims to alleviate data imbalance problem and sequential learning at the same time. Under the application of real air pollution data forecast, the proposed MCOS-ELM was compared with

retrained ELM and OS-ELM over accuracy and time measures. Experimental results show that MCOS-ELM has the highest efficiency and best accuracy for predicting the minority class (i.e., the most important but with fewest training samples) of air pollution level.

TABLE OF CONTENTS