Ecological Indicators 81 (2017) 362–372

Contents lists available at ScienceDirect

Ecological Indicators

journal homepage: www.elsevier.com/locate/ecolind

Water eutrophication evaluation based on semi-supervised classification: A case study in Three Gorges Reservoir

Di Wu[a,b,1], Huyong Yan[a,b,1], Mingsheng Shang[a], Kun Shan[a], Guoyin Wang[a,*]

[a] Chongqing Key Laboratory of Big Data and Intelligent Computing, Chongqing Institute of Green and Intelligent Technology, Chinese Academy of Sciences, Chongqing 400714, China

[b] University of Chinese Academy of Sciences, Beijing 100049, China

ARTICLE INFO

Keywords:
Semi-supervised classification
Eutrophication evaluation
Three Gorges Reservoir

ABSTRACT

Water eutrophication, which refers to the enrichment of nutrients to an aquatic environment, is one of the most challenging problems in water protection. Although many researchers have made attempts to solve the eutrophication problem, there is one issue that needs to be further discussed, i.e., how to establish a fast, low-cost, and accurate eutrophication evaluation model? For addressing this issue, this paper proposes a data-driven eutrophication evaluation model based on the semi-supervised classification. Concretely, a case study in Three Gorges Reservoir of China is carried out to demonstrate the validity of the proposed model. Experimental results clearly show that the proposed model has the advantages of high computational efficiency, high accuracy, and great ability of exploiting low-cost factors to assist or even replace high-cost factors in realizing the eutrophication evaluation. Moreover, we find that three low-cost factors, including *pH*, dissolved oxygen, and ammonia-nitrogen, are effective in achieving a better eutrophication evaluation for Three Gorges Reservoir based on the proposed model.

1. Introduction

Water is one of the most important resources for human survival and economic development (Li et al., 2016a; Li et al., 2015). Unfortunately, water, including groundwater and surface-water, is undergoing different degrees of deterioration all over the world because of the intensive human activity (Wu and Sun, 2016). For example, serious groundwater pollution in western China is induced by rapid urbanization and industrialization (Li, 2016; Li et al., 2016b), and the degraded water quality in the United States is caused by overfishing and increased aquaculture (Heisler et al., 2008). Water eutrophication, which refers to the enrichment of nutrients (nitrogen and phosphorus) to an aquatic environment, is one of the most challenging problems in water protection (Heisler et al., 2008). Similarly, due to the increased discharge of nutrients from industrialization, agricultural modernization, and urbanization, eutrophication is often reported and has attracted many attentions from both public and government (Du et al., 2011; Phillips et al., 2013). Nitrogen and phosphorus are necessary elements for plant growth. However, if they are input into water body more than necessary, the ecosystem will be changed, such as harmful algal blooms (HABs) and high levels of phytoplankton biomass, resulting in degradation of water quality (Qin et al., 2013). Eutrophication has posed a threat to the safety of resident drinking water, the natural ecological environment, and economic development (Heisler et al., 2008). For instance, cyanobacterial blooms resulted from eutrophication caused a crisis for the Wuxi drinking water in 2007 (Qin et al., 2010). Thus, we must adopt proper measures to control eutrophication (Yang et al., 2008).

Obviously, the premise of controlling the water eutrophication is to establish the appropriate methods or models to evaluate the trophic status of water body. So far, many researchers have made attempts to address this issue. In the environmental and ecological fields, there are many models, including Carlson trophic state index (Carlson, 1977), modified Carlson trophic state index (Aizaki, 1981), trophic state (Vollenweider et al., 1998), comprehensive nutrition state index (Xu et al., 2012), phytoplankton trophic index (Phillips et al., 2013), species diversity index (Spatharis and Tsirtsis, 2013), integrated methodology (Wu et al., 2013), eutrophication index (Fertig et al., 2014), etc. In the field of informatics, neural network (Kuo et al., 2007; Melesse et al., 2016; Singh et al., 2012; Yang et al., 2015), genetic algorithm (Song et al., 2012), fuzzy set theory (Giusti et al., 2011), support vector machine (SVM) (Huo et al., 2014), and rough set theory (Yan et al., 2016b)

* Corresponding author.
E-mail addresses: wudi@cigit.ac.cn, wanggy@ieee.org (G. Wang).
[1] These authors contributed equally to this work.

http://dx.doi.org/10.1016/j.ecolind.2017.06.004
Received 31 January 2017; Received in revised form 30 May 2017; Accepted 4 June 2017
1470-160X/

have been used in eutrophication evaluation. However, these methods or models still are not satisfactory in some cases because the eutrophication evaluation is nonlinear, multi-factors influenced, and complex in water ecological system (Ding and Wang, 2013). Additionally, due to the rapid development of automatic monitoring techniques (Dong et al., 2015), these methods or models are facing some new challenges as bellows:

- Due to the long time interval of traditional manual sampling, the collected monitoring data is small in most cases. Under such situation, the most existed eutrophication evaluation models were designed without considering the data processing ability. Thanks to the rapid development of automatic monitoring techniques, we can continuously collect massive monitoring data in 24 h. The massive monitoring data can effectively and accurately reflect the real conditions of water quality in time (Arienzo et al., 2015). Unfortunately, the explosive growth of monitoring data has brought some new challenges to the existed eutrophication evaluation models. For instance, How to effectively process the monitoring data with TB level is an important problem for the existed eutrophication evaluation methods or models.
- For most existed eutrophication evaluation models, total nitrogen (*TN*), total phosphorus (*TP*), Chlorophyll a (*Chl-a*), Secchi depth (*SD*), and Permanganate index (COD_{Mn}) are the key factors (Qin et al., 2013). Besides, some other factors, such as water temperature (*T*), dissolved oxygen (*DO*), *pH*, conductivity (*Cond.*), ammonia-nitrogen (NH_3-N), suspended solid (*SS*), position of sampling site (*site*), and season of sampling (*season*), are also connected with eutrophication (Giusti et al., 2011; Wu et al., 2013). However, the costs of collecting these factors are different because of their different monitoring principles. For example, *T, DO, pH, Cond.*, NH_3-N, and *SS*, can be collected easily by using the online sensors, while *TN, TP*, and COD_{Mn} are relatively difficult to collect because they demand the complicated pretreatment processes (A Xylem Brand, 2017). Thus, how to exploit the low-cost factors to assist or even replace the high-cost factors to realize the eutrophication evaluate is worth studying.
- In general, the existed eutrophication evaluation models rarely consider the ability of processing incomplete information. However, situation of missing key data happens frequently in the eutrophication evaluation because of the reasons of laboratory errors, instrument malfunctions, and even human errors. Therefore, it is necessary to take into account the ability of processing incomplete information for an eutrophication evaluation.

Classification, which relies heavily on the training instances with class labels, is an active research issue in data mining and machine learning communities (Cococcioni et al., 2012; Luo et al., 2015a; Su et al., 2009). Nevertheless, due to the technical support from experts as well as long time consumption of manual process for data labeling, it is difficult to obtain sufficient labeled data for supporting the classification tasks. Having a multitude of unlabeled data and few labeled ones is a common phenomenon in many practical applications (Zhou and Li, 2010). A successful and special methodology to tackle this problem is semi-supervised classification (SSC). SSC is highly effective in alleviating the shortage of labeled instances in classification tasks by exploiting the abundant unlabeled data (Triguero et al., 2015b). Thus, SSC has been widely and frequently used in several areas, including fault diagnosis, remote sensing monitoring, face recognition, *etc* (Schwenker and Trentin, 2014; Zhu, 2008). However, until now SSC has not been widely applied in eutrophication, and only few researchers have reported some preliminary studies on SSC applications in the fields of water supply (Herrera et al., 2010) and water quality retrieving (Wang et al., 2011).

Note that eutrophication evaluation actually is a classification task and the three challenges discussed above can be conquered by SSC. Thus, this paper innovatively proposes to use the SSC to achieve a powerful eutrophication evaluation model to overcome the three challenges discussed above. To the best of our knowledge, this paper is the first one to analyze eutrophication based on SSC and is different from the traditional environmental and ecological eutrophication analysis because it is completely data-driven. In order to illustrate the principle and usefulness of our proposal, a case study in Three Gorges Reservoir (TGR) of China is conducted in this paper. The experimental results clearly validate the fact that our proposal is a promising alternative method to achieve eutrophication evaluation with strong generalization ability.

The remainder of this paper is organized as follows: Section 2 introduces the materials and methods. Section 3 presents the proposed model. Section 4 provides and discusses the experimental results. Finally, Section 5 concludes this paper.

2. Materials and methods

2.1. Materials

2.1.1. Study area

TGR, which is created by the Three-Gorge Dam, sites at 29°16′–31°25′N, 106°–111°10′E, and has a surface area of 1080 km^2 (Zeng et al., 2006). The eutrophication in TGR has become the main environmental problem and attracts more and more attention from worldwide (Yan et al., 2016b). As shown in Fig. 1, the water level of TGR experiences two phases throughout the whole year because of flood management and water supply (Changjiang Maritime Safety Administration, 2017). In the ascending phase, the water level gradually increases from 145 *m* to 175 *m* and the flow speed becomes slow. The slow flow reduces the water exchange between the mainstream and the tributaries, resulting in the deposition of nutrients. Consequently, the eutrophication and even the HABs appeared in some tributaries (Yang et al., 2010). In the descending phase, the flow speed becomes fast, which intensifies the water exchange between the mainstream and the tributaries. As a result, water stability is reduced and the concentration of suspended silt is increased, which induces the frequent changes of water trophic state (Yang et al., 2010). Hence, the eutrophication evaluation of TGR should be high frequent for a better understanding on the water trophic state of TGR. Moreover, under such special hydrological conditions, the eutrophication evaluation of TGR also faces the three challenges discussed in Section 1. Therefore, we selected five typical tributaries in TGR as the study areas to validate the usefulness and effectiveness of our proposal, as shown in Fig. 2.

2.1.2. Field data

Data were obtained from several sampling sections located at five typical tributaries of TGR. The distributions of the tributaries are illustrated in Fig. 2. Sampling sections were visited during the period from

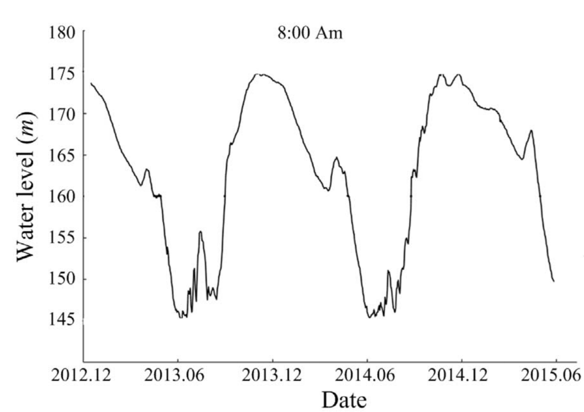

Fig. 1. The water level of TGR during two phases.