Ecological Indicators 107 (2019) 105664

Contents lists available at ScienceDirect

Ecological Indicators

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

Letter to the Editor

Predicting total phosphorus levels as indicators for shallow lake management

ARTICLE INFO

Keywords:
Algorithm's parameters
Data mining technique
Random forest
Variable selection

ABSTRACT

The discusser thanks the authors for investigating the ability of modified random forest algorithm to predicting total phosphorus levels as indicators for shallow lake management. The abilities of machine learning techniques such as optimization algorithms today have been well documented in engineering sciences. In this discussion, the discusser has tried to clarify the process of the paper of "Predicting total phosphorus levels as indicators for shallow lake management"(doi: https://doi.org/10.1016/j.ecolind.2018.09.002). The discusser would like to call attention to some important points, which may be taken into consideration by the authors and other potential researchers.

1. Introduction

In the original paper (Vitense et al., 2019), random forest algorithm has been used to methodology to build models for predicting these two TP classes using both remotely sensed watershed scale predictors and in-lake variables, and we performed recursive feature elimination to find the most parsimonious model. And the author used a modified random forest algorithm, RF++, to account for correlation among within-lake observations, and perform recursive feature elimination (RFE) to find the most parsimonious model in the context of correlated predictors, if used incorrectly this method can cause other researchers to Confused.

2. Results

The authors know that the accuracy and ability of random forest algorithm directly depend on this algorithm parameter. In random forest, hyper-parameters are used for increasing the speed of the model. Following are the scikit-learn's hyper-parameters. (1) n_jobs parameter: It provides the engine with details about the limit of processor for computational usage. If it has a "1" value, then this indicates that only a single processor can be run. On the contrary a "−1" value indicates that there is no restriction. (2) N_estimators parameter: t is the overall number of trees to be generated in the time period before the determination of max voting and averages for predictions. A larger number of figures increases reliability but it also affects the performance speed. (3) random_state parameter: It is used to convert the model's output to create a replicable result. If similar piece of training data, a definite value for random_state, and hyper-parameters are inserted in the model, then the output would also be identical. (4) min_sample_leaf parameter: It examines the lowest limit of a leaf for the split of internet nodes. (5) Max_features parameter: It takes the figure of maximum digit of features that are required to be used in each tree.

However, they did not provide how a random forest algorithm handled, and how find the optimal values of the control parameters mentioned in the study, this needs explanation. All the model parameters should be provided in the study by the authors. In this way, other researchers may also repeat the applications.

3. Conclusion

In the hydrological modeling the parameters of the models are very important (Moazenzadeh and Mohammadi, 2019; Jahani and Mohammadi, 2018; Aghelpour et al., 2019; Mohammadi, 2019a,b,c), so that the wrong choice of preprocessing can affect the modeling result. A study is valuable when it can be useful to other researchers, so the discusser invites the authors of the original paper to clarify their article and modify the items described, to make the original paper a useful reference for all researchers and by providing the random forest algorithm information, the original article would be clarified and the ambiguities mentioned clarified.

References

Aghelpour, P., Mohammadi, B., Biazar, S.M., 2019. Long-term monthly average temperature forecasting in some climate types of Iran, using the models SARIMA, SVR, and SVR-FA. Theor. Appl. Climatol. 1–10. https://doi.org/10.1007/s00704-019-02905-w.

Jahani, B., Mohammadi, B., 2018. A comparison between the application of empirical and ANN methods for estimation of daily global solar radiation in Iran. Theor. Appl. Climatol. 137 (1–2), 1257–1269. https://doi.org/10.1007/s00704-018-2666-3.

Moazenzadeh, R., Mohammadi, B., 2019. Assessment of bio-inspired metaheuristic optimisation algorithms for estimating soil temperature. Geoderma 353, 152–171. https://doi.org/10.1016/j.geoderma.2019.06.028.

Mohammadi, B., 2019a. Letter to the editor "Estimation of sodium adsorption ratio indicator using data mining methods: a case study in Urmia Lake basin, Iran" by Mohammad Taghi Sattari, Arya Farkhondeh, and John Patrick Abraham. Environ. Sci. Pollut. Res. 1–2. https://doi.org/10.1007/s11356-019-04368-y.

Mohammadi, B., 2019b. "Prediction of effective climate change indicators using statistical downscaling approach and impact assessment on pearl millet (Pennisetum glaucum L.) yield through genetic algorithm in Punjab, Pakistan" by Asmat Ullah, Nasrin Salehnia, Sohrab Kolsoumi, Ashfaq Ahmad, Tasneem Khaliq. Ecol. Indic. 101, 973–974. https://doi.org/10.1016/j.ecolind.2019.02.013.

Mohammadi, B., 2019c. Letter to the Editor "Design of an integrated climatic assessment in-

https://doi.org/10.1016/j.ecolind.2019.105664
Received 2 May 2019; Received in revised form 16 August 2019; Accepted 20 August 2019
Available online 27 August 2019
1470-160X/

dicator (ICAI) for wheat production: A case study in Jiangsu Province, China" by Xiangying Xu, Ping Gao, Xinkai Zhu, Wenshan Guo, Jinfeng Ding, Chunyan Li, Min Zhu, Xuanwei Wu. Ecol. Indic. 103, 493. https://doi.org/10.1016/j.ecolind.2019.04.055.

Vitense, K., Hanson, M.A., Herwig, B.R., Zimmer, K.D., Fieberg, J.R., 2019. Predicting total phosphorus levels as indicators for shallow lake management. Ecol. Indic. 96 (1), 278–287. https://doi.org/10.1016/j.ecolind.2018.09.002.

Babak Mohammadi

College of Hydrology and Water Resources, Hohai University, Nanjing 210098, China

E-mail addresses: Babakmohammadi@aol.com, Babak@hhu.edu.cn.