Geophysical Research Abstracts
Vol. 21, EGU2019-10493, 2019
EGU General Assembly 2019

Exploring the Effective Evaluation Indices of Self-Organizing Map for Clustering Regional Flood Inundation Map

Li-Chiu Chang (1), Wu-Han Wang (1), and Fi-John Chang (2)

(1) Department of Water Resources and Environmental Engineering, Tamkang University, New Taipei City, Taiwan (changlc@mail.tku.edu.tw), (2) Department of Bionenvironmental Systems Engineering, National Taiwan University, Taipei, Taiwan (changfj@ntu.edu.tw)

Today, Artificial Intelligence is one of popular issues with many research topics and practical applications. The relative AI issues on the study of water resource management or flood forecast have become one of important topics. The Self-organizing maps (SOM) have the topological ordering property that can categorize a large number of regional inundation maps into a meaningful topology. The main purpose of this study is to propose the methodology to automatically build the Self-organizing maps (SOM) on clustering the flood spatial distribution. There are three major problems on building the SOM model. The first one is the topological error, that is, the twisting of the map; the second one is to the selection of the number of epochs in two phases, ordering phase and convergent phase; the third one is to decide the optimal size.

This study proposes two training strategies of the SOM models to investigate the convergence of the SOM models. The first strategy, called plan1, is to train the network in the ordering phase until the weights of the neurons have no obvious change, then transfer to the convergent phase and continue training the neurons until the weights have no obvious change. The second strategy, called plan2, is to train the network in the ordering phase until the coverage rate of weights reaches 50%, then transfer to the convergent phase and continue training the same as the convergent phase of plan1. The study areas are Luermen Creek and Yenshui Creek located in Tainan City, Taiwan, and Kemaman River located in Terengganu, Malaysia. We use the flood simulation data of these three areas as the training data to build their own models. Through the different training strategy of plan1 and plan2, we can explore the influences of the ordering and convergent phases on building the SOM models. Through coverage rate, flip detector and five indices to compare the clustering results of the SOM clustering results. The coverage rate is defined as the difference of the cumulative distribution rates between maximum and minimum weights (neurons). The flip detector can check whether any two or more neurons flip each other weights or not and determine topological order correct or not.

The clustering results of these three cases show that the number of epochs can influence the coverage rate and effectively improve the clustering quality. Plan2 can get convergent clustering results while plan1 occurs flip in Luermen Creek and Kemaman River; that means plan2 is more suitable than plan1 on clustering the flood spatial distribution. Moreover, the coverage rates of 3×3 model are smaller than those of 4×4 and 5×5 models, about 5%-10% less. The coverage rates of 4×4 and 5×5 models are almost the same, so the smaller model should be enough neurons to describe the data, that is, the network of 4×4 is more suitable than other models. Hence, for choosing the size of topology map, the coverage rate is the great index to decide the optimal size.