Hydroinformatics in Hydrology, Hydrogeology and Water Resources (Proc. of Symposium JS.4 at the Joint IAHS & IAH Convention, Hyderabad, India, September 2009). IAHS Publ. 331, 2009, 224-235.

Data mining for hydrological time series analysis

RULIN OUYANG[1,2], LILIANG REN[1] & CHENGHU ZHOU[2]

1 *State Key Laboratory of Hydrology, Water Resources and Hydraulic Engineering, Hohai University, Nanjing 210098, China*
oyrl@hhu.edu.cn

2 *State Key Laboratory of Resources and Environmental Information System, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China*

Abstract The rapid development of data mining supplies a new way for hydrological information analysis and mining, which use artificial intelligence methodology. Applying data mining theory and technology, we analysed hydrological daily discharge time series of Shaligunlanke station in the Tarim River basin in China for the years 1961–2000. First, according to the four monthly statistics, mean monthly discharge, monthly maximum discharge, monthly amplitude and monthly standard deviation, *K*-mean clustering was used to segment the annual process of the daily discharge. The clustering result showed that the annual process of the daily discharge can be divided into five segments: snowmelt period I (April), snowmelt period II (May), rainfall period I (June–August), rainfall period II (September) and dry period (October–December and January–March). Secondly, dynamic time warping (DTW), which is a different distance metric method from the traditional Euclidian distance metric, was used to search for similarity in the discharge process. Based on the similarity matrix, the similar discharge processes can be mined in each period. Finally, the physical causes of the similar processes identified in the previous steps were analysed. It was found that the discharge had a close relationship with the temperature and the precipitation, and the discharge processes were more similar under the same climatic conditions. Our study shows that data mining is a feasible efficient approach to discovering the hidden information in the historical hydrological data and to mining the implicative laws under the hydrological process.

Key words data mining; hydrological time series; dynamic time warping; similarity search