

一种基于信息熵的数据预处理技术

阮祥超, 万定生

(河海大学 计算机与信息学院, 江苏 南京 211100)

摘要: 神经网络是数据挖掘领域中的重要研究课题.针对水文数据不规则和非线性的数据特征, 提出了一种基于信息熵的数据预处理方法, 使用了基于改进的 BP 神经网络模型. 该模型采用计算范围熵的熵值的方法, 遍历查找替换输入变量, 使得输入数据集的变量间线性关系减弱, 非线性关系增加, 熵值增大, 达到提高水文预测精度的目的.在实际的预测中基于相关信息熵的改进神经网络模型的预测效果优于传统预测模型.

关键词: 数据挖掘; 神经网络; 相关信息熵; 水文预测

A Data Preprocessing Technique Based on Information Entropy

RUAN Xiang-chao, WAN Ding-sheng

(College of Computer and Information, HoHai University, Nanjing 211100, China)

Abstract: Neural network is an important research topic in the field of data mining. This paper proposing a data preprocessing method based on information entropy which is directed to the irregular and non-linear data characteristics of hydrological data, and uses an improved model of BP neural network. The model uses the method of calculating the entropy of the range entropy, traversing the input variables, searching and replacing them, making the linear relationship between the variables of the input data set weakened. By this way, the non-linear relationship increasing, and the value of the information entropy increasing, so that it can achieving the purpose of improving the accuracy of hydrological prediction. In the actual prediction, the improved model of neural based on the relevant information entropy is superior to the traditional prediction model.

Key words: data mining; neural network; information entropy; hydrological forecasting

作者简介:

阮祥超 女, (1994-), 硕士研究生.研究方向为信息处理与信息系统.

E-mail:xchruan@hhu.edu.cn.

万定生 男, (1963-), 教授.研究方向为信息处理与信息系统.