基于改进遗传算法的不完整大数据填充挖掘算法

郑 斌
(福建江夏学院 电子信息科学学院,福建 福州 350108)

摘 要:传统的大数据填充挖掘过程存在耗时高、效率低的缺陷,对此提出基于改进遗传算法的不完整大数据填充挖掘算法.该算法先融合于BP人工网络神经算法随机产生不完整大数据初始网络权值,并对该组权值进行训练,设计一个不完整大数据基因矩阵,利用遗传算法以该基因矩阵为依据计算出不完整大数据适应值函数,并同时进行编码、杂交、变异等遗传操作,在此基础上进行不完整大数据信息的全局搜索,以其搜索的结果为核心将大数据划分为完整与不完整数据,采用信息论中熵值的概念利用同一类完整数据信息的属性值对缺失值进行填充.实验仿真证明,基于改进遗传算法的不完整大数据填充挖掘方法能实现对不完整大数据的挖掘,提高缺失数据的填充精度.

关键词:不完整大数据;数据填充;遗传神经网络
中图分类号:TP311.13 **文献标识码**:A **文章编号**:1000-7180(2016)02-0096-04

Incomplete Data Filling Mining Algorithm Based on the Improved Genetic Algorithm

ZHENG Bin
(College of Electronics and Information Science, Fujian Jiangxia University, Fuzhou 350108, China)

Abstract: Traditional big data high filling mining process is time-consuming, the defects of low efficiency. Based on improved genetic algorithm of incomplete data filling mining algorithm. This algorithm fusion in the BP artificial neural network algorithm first randomly generated incomplete data the initial network weights, and to train the weights of the group to design a incomplete data gene matrix, by using the genetic algorithm based on the gene matrix to calculate the incomplete data adaptive value function, and at the same time coding, hybridization and mutation genetic operation, on the basis of incomplete big data information of global search, with its search results as the core will be divided into complete and incomplete data, using the concept of entropy in information theory, and the same kind of complete data information of attribute values to fill the missing value. The experimental simulation proves that the incomplete big data based on improved genetic algorithm of incomplete filling mining method can realize data mining, improve the filling precision of the missing data.

Key words: incomplete data; Data filled; Genetic neural network

作者简介:
郑 斌 女,(1977-),硕士,讲师.研究方向为信息处理和数据分析.E-mail:frostzb@163.com