

基于关联规则的医疗大数据挖掘算法

岳根霞

(山西医科大学汾阳学院, 山西 汾阳 032200)

摘要: 针对当前医疗大数据挖掘过程中医疗大数据挖掘效果差的问题, 提出一种基于关联规则的医疗大数据挖掘算法, 通过对医疗大数据的频繁项集进行查找, 产生强关联规则, 并对医疗大数据集的平均值和标准差的计算, 实现对关联规则的调整, 确定医疗大数据属性, 构建合适的适应度函数, 根据适应度函数值进行选择、交叉、变异操作得到约简的分类属性组合, 对约简的分类属性信息熵进行计算, 并通过计算期望信息得到分类属性的信息增益, 建立决策树挖掘模型, 实现医疗大数据挖掘. 实验结果表明, 所提算法能够有效提高医疗大数据预处理的水平, 数据挖掘的效果较好.

关键词: 关联规则; 医疗; 大数据; 挖掘; 算法

Big Data Mining Algorithm Based on Association Rules

YUE Gen-xia

(Fenyang College of Shanxi Medical University, Fenyang 032200, China)

Abstract: In view of the current medical data mining process, the medical data mining problems, proposes a mining algorithm based on association rules of medical data, searching through frequent itemsets of medical data, generate strong association rules, the calculated average value and standard deviation of large medical data sets, implementation the association rules adjustment, determine the medical data attributes, construct the appropriate fitness function, according to the classification attribute combination value of fitness function selection, crossover and mutation reduction, calculation of information entropy on classification attribute reduction, classification and attribute information gain by calculating the expected information, the establishment of decision tree mining model, implementation of medical data mining. The experimental results show that the proposed algorithm can effectively improve the level of medical data preprocessing, and the effect of data mining is better.

Key words: association rules; medical; big data; mining; algorithms

作者简介:

岳根霞 女, (1977-), 硕士, 副教授. 研究方向为大数据、数据挖掘. E-mail: yuegenxia@163.com