计算机科学与探索 2007 1 (3): 314-324 ISSN: 1673-9418 CN: 11-5602/TP

学术研究

基于相对等待时间的代价敏感决策树

袁鼎荣, 张师超+, 朱晓峰, 张晨

广西师范大学 计算机科学与信息工程学院,广西 桂林 541004

收稿日期 修回日期 网络版发布日期 2007-9-20 接受日期

摘要 首先引入相对等待时间代价, 将它与测试代价一起称为有形代价, 利用单位有形代价中无形代价（即误分类代价）降低最多的原则选择分裂属性;然后结合序列测试策略和批量测试策略建立相对等待时间代价敏感决策树。实验结果显示, 该方法无论在误分类代价的减少量上还是所需有形代价的数量上都优于存在的算法, 并且实验地分析了建立代价敏感决策树考虑相对等待时间是必要的。

关键词 代价敏感 决策树 分裂属性 等待时间 时间敏感

分类号

Cost-sensitive decision trees based on relative waiting time

YUAN Dingrong,ZHANG Shichao+,ZHU Xiaofeng,ZHANG Chen

College of Computer Science and Technology,Guangxi Normal University,Guilin,Guangxi 541004,China

Abstract

The paper defines a ‘relative waiting time cost’,and combines it with ‘test cost’ as ‘tangible cost’. Then a principle is presented for selecting splitting attributes,in which those attributes that maximally decrease intangible cost （i.e.,‘misclassification cost’） in an unit of tangible cost,will be firstly selected as splitting attributes. Furthermore,cost-sensitive decision trees with relative waiting time cost is proposed by combining the sequential test strategy with batch test strategy. Experimental results demonstrate that the method outperforms the existing methods not only at the decrease of intangible cost,but also at the quantity of the tangible cost. On the other hand,the results show that it is necessary to consider the relative waiting time cost when building cost-sensitive decision trees.

Key words cost-sensitive decision trees splitting attributes waiting time time-sensitive

DOI:

通讯作者 袁鼎荣 E-mail:zhangsc@mailbox.gxnu.edu.cn