

本期目录 | 下期目录 | 过刊浏览 | 高级检索

[打印本页] [关闭]

博士论文

基于概率论的隐私保持分类挖掘

李光, 王亚东, 苏小红

(哈尔滨工业大学计算机科学与工程系, 哈尔滨 150001)

摘要: 在现有的基于数据扰动的隐私保持分类挖掘算法中, 扰动数据和原始数据相关联, 对隐私数据的保护并不完善, 且扰动算法和分类算法耦合度高, 不适合在实际中使用。为此, 提出一种基于概率论的隐私保持分类挖掘算法。扰动后可得到一组与原始数据独立同分布的数据, 使扰动数据和原始数据不再相互关联, 各种分类算法也可直接应用于扰动后的数据。

关键词: 数据挖掘 隐私保持 数据扰动 随机噪声 决策树

Privacy-preserving Classification Mining Based on Probability Theory

LI Guang, WANG Ya-dong, SU Xiao-hong

(Department of Computer Science and Engineering, Harbin Institute of Technology, Harbin 150001, China)

Abstract: In the existed privacy-preserving classification mining methods based on data perturbation, the privacy data is not protected perfectly because the perturbed data and the original data have been related. The classification algorithm and the data perturbation algorithm have high coupling It is not easy to use these methods in practice. To solve these problems, it proposes a privacy-preserving classification mining algorithm based on probability theory. The perturbed data is independent from the original data and they have the same distribution. This proposed method overcomes the shortcomings of others. The perturbed data is no relation with the original data and the classification methods can be used on the perturbed data directly.

Keywords: data mining privacy protection data perturbation random noise decision tree

收稿日期 2011-07-25 修回日期 网络版发布日期 2012-02-05

DOI: 10.3969/j.issn.1000-3428.2012.03.005

基金项目:


国家“863”计划基金资助项目(2007AA02Z329)

通讯作者:

作者简介: 李光(1982—), 男, 博士研究生, 主研方向: 数据挖掘, 生物信息学; 王亚东、苏小红, 教授、博士生导师

通讯作者E-mail: hit6006@126.com

参考文献:

- [1] Agrawal R, Srikant R. Privacy-preserving Data Mining[J].ACM SIGMOD Record.2000, 29 (2): 439-450 
- [3] Slava K, Lior R, Yuval E, et al. Efficient Multidimensional Suppression for K-anonymity

扩展功能

本文信息

- ▶ Supporting info
- ▶ PDF(230KB)
- ▶ [HTML] 下载
- ▶ 参考文献[PDF]
- ▶ 参考文献

服务与反馈

- ▶ 把本文推荐给朋友
- ▶ 加入我的书架
- ▶ 加入引用管理器
- ▶ 引用本文
- ▶ Email Alert
- ▶ 文章反馈
- ▶ 浏览反馈信息

本文关键词相关文章


- ▶ 数据挖掘
- ▶ 隐私保持
- ▶ 数据扰动
- ▶ 随机噪声
- ▶ 决策树

本文作者相关文章

- ▶ 李光
- ▶ 王亚东
- ▶ 苏小红

PubMed

- ▶ Article by Li, G.
- ▶ Article by Wang, E. D.
- ▶ Article by Su, X. G.

[5] Huang Zhengli.[J].Du Wenliang, Chen Biao. Deriving Private Information from Randomized Data[C]//Proceedings of ACM' s Special Interest Group on Management of Data. Baltimore, Maryland, USA: [s. n.].2005,:- 

本刊中的类似文章

1. 朱小栋, 樊重俊, 杨坚争.面向机场场区管理的数据挖掘系统[J]. 计算机工程, 2012,38(3): 224-227
2. 秦丽君, 罗雄飞.基于动态项集计数的加权频繁项集算法[J]. 计算机工程, 2012,38(3): 31-33
3. 刘佳新, 严书亭, 任家东.缩减投影数据库规模的增量式序列模式算法[J]. 计算机工程, 2012,38(3): 28-30
4. 潘冬生, 章昭辉, 代秀娟, 杨娟.基于层次梯度分析的协同数据挖掘算法[J]. 计算机工程, 2012,38(2): 72-74
5. 刘芳.基于图和双向搜索的频繁项集挖掘算法[J]. 计算机工程, 2012,38(01): 59-61
6. 姜丽莉, 孟凡荣, 周勇.多值属性关联规则挖掘的 Q-Apriori算法?[J]. 计算机工程, 2011,37(9): 81-83
7. 刘文远, 张亮, 孙德杰, 陈子军.改进的SOD孤立点检测算法[J]. 计算机工程, 2011,37(9): 93-94,97
8. 汪海锐, 李伟.基于关联规则的决策树算法[J]. 计算机工程, 2011,37(9): 104-106,109
9. 花红娟, 张健, 陈少华.基于频繁模式树的约束最大频繁项集挖掘算法[J]. 计算机工程, 2011,37(9): 78-80
10. 金聪, 谈华永, 王晓燕.基于博弈理论的计算机病毒传播模型[J]. 计算机工程, 2011,37(9): 155-156

文章评论

反馈人	<input type="text"/>	邮箱地址	<input type="text"/>
反馈标题	<input type="text"/>	验证码	<input type="text" value="1960"/>
<input type="text"/>			