

数据库、信号与信息处理

## 大规模数据集的增量式关联规则挖掘

张根香, 陈海山

厦门大学 软件学院, 福建 厦门 361005

收稿日期 2009-4-7 修回日期 2009-6-25 网络版发布日期 2009-10-10 接受日期

**摘要** 商业活动和工程实践中通常会积累一些大规模的携带重要信息的数据, 由于这种数据集经常有更新且数据量较大, 在对它们进行增量式关联规则挖掘时, 若采用基于传统的Apriori算法进行计算, 一方面难以取得较好的效率; 另一方面支持度设置过低会产生大量的冗余规则, 设置过高则会把一些支持度不高但有用的规则过滤掉而导致算法对这些新规则感应迟钝。因此, 借助遗传算法的相关机理, 同时结合自然界的免疫进化理论及相关仿生机制, 提出一种IOGA (Immune Optimization based Genetic Algorithm, 基于免疫优化的遗传算法) 增量式关联规则挖掘方法。通过实验表明, 该方法应用于大规模数据集的增量式关联规则挖掘时, 可以及时地感知规则的变更并发现有用的规则, 减少了冗余规则的产生, 同时挖掘效率也有明显提高。

**关键词** [免疫优化](#) [遗传算法](#) [关联规则](#) [增量式挖掘](#)

分类号 [TP31](#)

## Incremental association rules mining for large data set

ZHANG Gen-xiang, CHEN Hai-shan

Software School of Xiamen University, Xiamen, Fujian 361005, China

### Abstract

Business activity and engineering practice always accumulate large dataset with important information. But because of the dataset's largeness and frequent updating, if the Apriori based algorithm is applied to incremental association rules mining, it is not only inefficient, but also many redundant rules will be produced with low minimum support while some interesting rules will be lost with high minimum support, which leads to the algorithm's blunt perception to those rules. So, following genetic principle, and combining with natural immune involution theory and relative bionic mechanism, this paper proposes an IOGA (Immune Optimization based Genetic Algorithm) approach for incremental association rules mining. Experiment demonstrates the proposed method's effectiveness and presents its good performance in perceiving rules' subtle change, reducing redundant rules and finding interesting rules.

**Key words** [immune optimization](#) [genetic algorithm](#) [association rules](#) [incremental mining](#)

DOI: 10.3778/j.issn.1002-8331.2009.29.036

通讯作者 张根香 [zhanggenxiang2007@yahoo.cn](mailto:zhanggenxiang2007@yahoo.cn)

### 扩展功能

#### 本文信息

- ▶ [Supporting info](#)
- ▶ [PDF\(949KB\)](#)
- ▶ [\[HTML全文\]\(0KB\)](#)
- ▶ [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](#)
- ▶ [复制索引](#)
- ▶ [Email Alert](#)
- ▶ [文章反馈](#)
- ▶ [浏览反馈信息](#)

#### 相关信息

- ▶ 本刊中 [包含“免疫优化”的相关文章](#)
- ▶ 本文作者相关文章

- [张根香](#)
- [陈海山](#)