

# 一种有效的多关系贝叶斯分类算法

邓左祥<sup>1, 2,3</sup>, 涂芳<sup>4</sup>

(1 广西科技大学 计算机科学与通信工程学院, 广西 柳州 545006;

2 广西科技大学 智能计算与分布式信息处理重点实验室培育基地, 广西 柳州 545006;

3 上海交通大学 电子信息与电气工程学院, 上海 200240;

4 上汽通用五菱汽车股份有限公司 技术中心产品认证科, 广西 柳州 545007)

摘要: 传统数据挖掘算法在处理多关系时, 需要物理连接, 因此存在效率不高的问题. 为了解决这一问题, 研究多关系数据挖掘中的分类问题, 提出一种有效的多关系贝叶斯分类算法 EMBC. EMBC 算法的目标是提高分类的准确率, 并且降低运行时间. EMBC 算法利用元组 ID 传播的思想, 结合朴素贝叶斯分类算法, 可以直接在多关系上进行分类. 实验表明, EMBC 算法提高分类的准确率, 并且显著降低运行时间.

关键词: 数据挖掘; 多关系; 分类; 元组 ID 传播; 贝叶斯

## Efficient Multi-relational Bayesian Classification Algorithm

DENG Zuo-xiang<sup>1, 2,3</sup>, TU Fang<sup>4</sup>

(1 School of Computer Science and Communication Engineering, Guangxi University of Science and Technology, Liuzhou 545006, China; 2 Key Laboratory of Intelligent Computing and

Distributed Information Processing, Guangxi University of Science and Technology, Liuzhou 545006, China; 3 School of Electronic Information and Electrical Engineering, Shanghai Jiao Tong University, Shanghai 200240, China; 4 Technical Department Center Test & Validation Section,

SAIC GM Wuling Automobile CO., Ltd., Liuzhou 545007, China)

Abstract: While dealing with multi-relation, traditional data mining algorithms used physical join, thus it had the problem of low efficiency. In order to solve this problem, the problem of classification in multi-relational data mining was investigated, and an efficient multi-relational Bayesian classification algorithm called EMBC was proposed. EMBC aims at increasing the accuracy of classification, and decreasing running time. By taking advantage of tuple ID propagation approach, and combined with naive Bayesian classification algorithm, EMBC can directly classify in multi-relation. Performance results demonstrate that, EMBC increases the accuracy of classification, and significantly decreases running time.

Key words: data mining; multi-relation; classification; tuple ID propagation; Bayesian

作者简介:

邓左祥 男, (1983-), 博士. 研究方向为数据挖掘、自组织网络. E-mail: dengzuoxiang@126.com.

涂芳 女, (1983-), 工程师. 研究方向为数据挖掘.