

一种有效的多关系聚类算法

邓左祥^{1, 2}, 李春贵¹

(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 上海交通大学 电子信息与电气工程学院, 上海 200240)

摘要: 研究多关系数据挖掘的聚类问题, 提出一种有效的多关系聚类算法 EMC. EMC 算法的目标是提高聚类的准确率, 并且降低运行时间. EMC 算法首先利用元组 ID 传播的思想, 计算两个对象之间的相似度, 接着利用 K 中心点聚类算法, 将对象划分成簇. 实验表明, EMC 算法显著降低运行时间, 并且提高聚类的准确率.

关键词: 多关系数据挖掘; 聚类; 元组 ID 传播; 相似度; K 中心点聚类算法

Efficient Multi-Relational Clustering Algorithm

DENG Zuo-xiang^{1, 2}, LI Chun-gui¹

(1 School of Computer Science and Communication Engineering, Guangxi University of Science and Technology, Liuzhou 545006, China; 2 School of Electronic Information and Electrical Engineering, Shanghai Jiao Tong University, Shanghai 200240, China)

Abstract: The problem of clustering in multi-relation data mining was investigated, and an efficient multi-relational clustering algorithm called EMC was proposed. EMC aims at increasing the accuracy of clustering, and decreasing running time. First, EMC computed the similarity between two objects by taking advantage of tuple ID propagation approach. Then, EMC clustered the objects by K-medoids clustering algorithm. Performance results demonstrate that, EMC significantly decreases running time, and increases the accuracy of clustering.

Key words: multi-relational data mining; clustering; tuple ID propagation; similarity; K-medoids clustering algorithm

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.

李春贵 男, (1968-), 博士, 教授. 研究方向为数据挖掘、人工智能.