

KHAI THÁC DỮ LIỆU - TÌM TẬP LUẬT KẾT HỢP THEO MỤC ĐÍCH NGƯỜI DÙNG

Lê Hoài Bắc, Phạm Thị Bạch Huệ

Khoa CNTT, Trường Đại học Khoa học Tự Nhiên - ĐHQG tp.HCM

Tóm tắt:

Đối với bài toán khai thác dữ liệu, luật kết hợp được xem là cách biểu diễn tri thức đơn giản và dễ hiểu. Hiện nay đã có nhiều thuật toán tìm luật kết hợp từ cơ sở dữ liệu lớn rất hiệu quả. Các thuật toán này thường cho ra một tập hợp rất lớn các luật kết hợp gồm cả những luật vô ích. Việc tìm ra luật nào thực sự cần thiết cho người dùng trong tập hợp rất lớn các luật là bài toán được đặt ra. Ta gọi đây là bài toán tìm tập luật kết hợp theo mục đích người dùng (Interesting rules - IR).

Hiện tại đã có nhiều tác giả quan tâm giải quyết bài toán này. Tuy nhiên, các cách tiếp cận còn rời rạc. Bài viết này mục đích tổng hợp một cách có hệ thống các phương pháp tìm tập luật IR. Việc cài đặt thử nghiệm và so sánh kết quả đạt được của một số phương pháp giúp người dùng ứng dụng luật kết hợp hiệu quả hơn trong công việc của họ.

DATA MINING - FINDING INTERESTING RULES FROM LARGE DATABASES

Le Hoai Bac, Pham Thi Bach Hue

Department of Information Technology, University of Natural Sciences -
VNU.HCM

Abstract:

Association rule is considered to be the simple and easy understanding way of knowledge expression. Efficient algorithms exist for discovering association rules from large collections of data. These algorithms often output a large set of association rules, including the useless ones. Browsing the rule set and finding interesting rules from it can be quite difficult for the user. The problem here is finding Interesting Rules (IR) - rules are really useful for the user - from large set of discovered association rules or directly from mining databases.

There are a lot of researchers have solved this problem by one way or another. However, these approaches are still discrete. This paper summarizes systematically some interesting approaches of finding IR set. The implementation and the comparison the efficiency of the concerned approaches help the user to use the association rules easier.