

MỘT CÁCH TIẾP CẬN TÌM TẬP PHỔ BIẾN DỰA TRÊN GIÀN TRONG KHAI PHÁ LUẬT KẾT HỢP

AN APPROACH TO BUILDING THE FREQUENCY SET BASED ON LATTICE IN MINING ASSOCIATION RULES

Tác giả: Lương Văn Nghĩa ; Lê Văn Sơn ; Huỳnh Triệu Vỹ

Tóm tắt bằng tiếng Việt:

Khai phá luật kết hợp trong các cơ sở dữ liệu giao dịch lớn là bài toán đã được nhiều người quan tâm nghiên cứu. Bài toán khai phá luật kết hợp thường được thực hiện qua hai bước. Trong đó, bước đầu tiên là tìm tập phổ biến và bước thứ hai tìm các luật kết hợp dựa trên tập phổ biến tìm được. Hiện đã có rất nhiều thuật toán tìm tập phổ biến và thuật toán đề xuất sinh giàn từ quan hệ nhị phân, tuy nhiên các thuật toán này có độ phức tạp rất lớn. Trong bài báo này chúng tôi giới thiệu một kỹ thuật tìm tập phổ biến dựa trên giàn có độ phức tạp đa thức. Ưu điểm của cách tiếp cận này là bỏ qua giai đoạn tìm tập ứng viên như trong thuật toán Apriori mà tìm trực tiếp tập phổ biến.

Từ khóa: Luật kết hợp; Tập phổ biến; Giàn; Lược đồ Hasse; Thuật toán Apriori.

Tóm tắt bằng tiếng Anh:

In recently years, the Discovery of Association Rule on the transaction of large databases has been the most interesting problem in research. The problem of mining association rule is usually performed through two steps. The frequency set is found in first step, and building the association rule based on the previous result of frequency set is second step. In fact, we had many algorithms to find the frequency set and to propose for generating lattices from binary relationships. However, those algorithms still have been a big complexity. In this paper, we introduce a technique to find frequency set based on the lattice in which there has been the complexity of polynomial. The advantage of this technique not only ignores the stage of finding candidate set as the Apriori algorithm, but also builds the frequency set immediately.

Key words: Association Rule; frequency set; lattice; Hasse Diagram; Apriori Algorithm.