

Các phương pháp khai thác tập có thể xóa trên dữ liệu tăng cường

NGUYỄN THỊ HOÀI LINH *

Trường Đại học Kinh tế - Tài chính TP.HCM
Nhận bài: 29/09/2021 - Duyệt đăng: 15/11/2021

(*) Liên hệ: linhnh@uef.edu.vn Tel: 0945.110.590

Tóm tắt:

Khai thác tập có thể xóa là một trong những lĩnh vực thú vị trong khai thác tập phổ biến. Phân tích dữ liệu tăng cường trở nên quan trọng hơn vì dữ liệu thú vị liên tục được tích lũy trong các lĩnh vực ứng dụng khác nhau bao gồm cả các lĩnh vực công nghiệp. Bài viết này cung cấp một cuộc khảo sát về các thuật toán khai thác tập có thể xóa (EI) hiện có trên cơ sở dữ liệu tăng cường. Thuật toán khai thác tập có thể xóa trên dữ liệu tăng cường được trình bày lần đầu vào năm 2016 và IWEI là thuật toán đầu tiên khai phá các tập có thể xóa xem xét các điều kiện có trọng số trong cơ sở dữ liệu tăng cường. Kể từ đó, một số thuật toán, chẳng hạn như FUP-ERASABLE, LINE và WEIL đã được đề xuất nhằm khai thác tập có thể xóa trên dữ liệu tăng cường. Trong nghiên cứu này, các thuật toán IWEI, FUP-ERASABLE, LINE và WEIL được mô tả và phân tích các điểm mạnh và nhược điểm của từng thuật toán.

Từ khóa: Khai thác dữ liệu, tập có thể xóa, khai thác tăng cường, điều kiện có trọng số, tập có thể xóa có trọng số.

Abstract:

Erasable itemset mining is one of the interesting areas in frequent itemset mining. Analyzing incremental data becomes more important because interesting data are continually accumulated in various application fields including industrial areas. This article provides a survey of the available erasable itemset (EI) mining on incremental databases. Incremental mining algorithm for erasable itemsets was first presented in 2016 and IWEI is the first algorithm for discovering erasable itemsets considering weight conditions in incremental databases. Since then, a number of algorithms, such as FUP-ERASABLE, LINE and WEIL have been proposed mining EI on incremental data. In this study, the IWEI, FUP-ERASABLE, LINE and WEIL algorithms are described and analyze the strengths and weaknesses of each algorithm.

Keywords: Data mining, erasable itemsets, incremental mining, weighted conditions, weighted erasable itemsets.