

BẢN NHẬN XÉT KHÓA LUẬN TỐT NGHIỆP

(HƯỚNG NGHIÊN CỨU)

*Tên đề tài: KHAI THÁC LUẬT CÓ THỨ TỰ TRONG CƠ SỞ DỮ LIỆU CHUỖI**Sinh viên thực hiện: Lê Thành Nam – 18127158**Lê Hồng Quang – 18127190**Giảng viên hướng dẫn: Lê Hoài Bắc***1. Chủ đề và ý tưởng nghiên cứu:**

Khai thác tập luật là một công việc khai thác dữ liệu phổ biến để phân tích một chuỗi các sự kiện hoặc ký hiệu. Nó bao gồm việc xác định các chuỗi sự kiện thường xuyên xuất hiện trong một chuỗi và sau đó kết hợp chúng để có được các tập luật thể hiện mối quan hệ chặt chẽ giữa các sự kiện. Nhưng một vấn đề quan trọng là mỗi luật yêu cầu một thứ tự nghiêm ngặt của các sự kiện. Kết quả là, các luật tương tự được xử lý khác nhau, mặc dù chúng trong thực tế thường mô tả một tình huống giống nhau. Để tìm một bộ luật nhỏ hơn, tổng quát hơn và có thể thay thế nhiều tập luật, khóa luận này giới thiệu một loại tập luật mới được gọi là tập luật có thứ tự một phần, trong đó các sự kiện trong luật được sắp xếp một phần. Để tìm một cách hiệu quả tất cả các luật này trong một chuỗi, một thuật toán hiệu quả có tên là POERM được trình bày và đánh giá thử nghiệm trên một số tập dữ liệu chuẩn cho thấy POERM có hiệu suất tốt

2. Phương pháp nghiên cứu:

Ý tưởng của thuật toán là sẽ bắt đầu tìm thành phần của luật $X \rightarrow Y$ theo thứ tự từ trái sang phải. Thuật toán POERM sử dụng định lý cắt giảm để xem xét tính hợp lệ của các luật giảm không gian tìm kiếm luật gốc (với m event set sẽ sinh ra $(2m - 1) \times (2m - 1)$ luật). Thuật toán xây dựng một bộ POERs chứa các luật $X \rightarrow Y$ hợp lệ. Đầu tiên, thuật toán sẽ tìm ra các chuỗi 1-event thỏa định nghĩa cắt giảm và có khả năng là tiền tố X của luật $X \rightarrow Y$ đồng thời lưu lại các thời điểm X diễn ra. Sau đó thuật toán tiến hành mở rộng chuỗi X thành các chuỗi 2-event từ chuỗi X 1-event cũ và cứ tiếp tục mở rộng đến khi không mở thêm được nữa và cũng lưu lại thời điểm của các event này. Tiếp đến, thuật toán tìm các hậu tố Y có thể là hậu tố của luật $X \rightarrow Y$ với tiền tố X được lưu trước đó cùng các thời điểm của chúng. Sau đó thuật

Giảng viên hướng dẫn - 1

toán sẽ cắt giảm hậu tố Y theo định nghĩa cắt giảm. Tiếp đến, hậu tố Y cũng sẽ mở rộng như tiền tố X. Cuối cùng, thuật toán ghép tiền tố và hậu tố thành $X \rightarrow Y$ và cắt giảm theo định lý cắt giảm để cho các bộ POERs hợp lệ. Thời điểm của các bước trên đều được lưu lại. Bên cạnh đó, thuật toán MiningXEventSet được sử dụng để hỗ trợ việc cắt giảm các chuỗi tiền tố X cũng như mở rộng chúng

3. Đóng góp Khoa học và thực tiễn:

Từ kết quả nghiên cứu của Philippe Fournier-Viger và cộng sự, khóa luận đã nghiên cứu các loại luật mới được đề xuất gọi là các luật có thứ tự một phần, trong đó các sự kiện trong luật được sắp xếp một phần, dùng để tìm các tập luật tổng quát hơn. Để tìm một cách hiệu quả tất cả các luật này trong một chuỗi, một thuật toán hiệu quả có tên là POERM – Khai thác tập luật có thứ tự một phần đã được nghiên cứu. Khóa luận đã cải tiến về mặt cấu trúc dữ liệu trong thuật toán để cho hiệu suất tốt hơn. Một số đánh giá thực nghiệm trên một số tập dữ liệu cho thấy thuật toán POERM có hiệu suất tốt, đồng thời khảo sát ảnh hưởng của các tham số đến thời gian thực thi và lượng bộ nhớ đã sử dụng khi thực thi thuật toán.

4. Quá trình thực hiện:

Sinh viên đã có nhiều nỗ lực để hoàn thành khóa luận. Cần chủ động hơn trong các công việc của mình

5. Báo cáo viết:

Nêu rõ được nội dung của vấn đề cần giải quyết

6. Trình bày trước hội đồng:

7. Công bố khoa học/ ứng dụng thực tế: không có

Đánh giá xếp loại: Khá

TP.HCM, ngày 22 tháng 8 năm 2022

Giảng viên hướng dẫn



Lê Hoài Bắc

Giảng viên hướng dẫn - 2