

Tên LATS: NÂNG CAO HIỆU QUẢ TÍNH TOÁN TÌM TẬP PHỔ BIẾN THEO HƯỚNG
TIẾP CẬN TẬP ĐÓNG

Chuyên ngành: Khoa học Máy tính

Mã số: 62 48 01 01

Tên NCS: Nguyễn Thành Trung

Khóa đào tạo: 3

Người hướng dẫn: PGS.TS. Nguyễn Phi Khứ

PGS.TS. Nguyễn Đình Thuận

Trường: Đại học Công nghệ Thông tin – ĐHQG-HCM

Những đóng góp của LATS:

Chương 2 đề xuất một cấu trúc trung gian gọi là tập kiến tạo P gồm các mẫu kiến tạo từ đó suy ra các tập mục đóng, tập phổ biến. Thuật toán xây dựng tập kiến tạo theo hướng tăng cường các giao tác là ConPatSet và hàm IncPatSet để cập nhật tập kiến tạo P. Khi tiết giảm hay loại bỏ các giao tác từ tập giao tác, hàm DesPatSet được sử dụng để cập nhật P.

Với CSDL gồm n giao tác, m hạng mục, độ phức tạp của thuật toán ConPatSet và IncPatSet để tạo lập nên P là $O(m \times n \times k^2)$, $k = |P|$.

Việc lưu trữ dữ liệu giao tác theo các dãy bit có ưu thế đáng kể. Trước hết làm giảm thiểu lượng bộ nhớ trong và ngoài. Tính toán trên dữ liệu bit nhanh hơn rất nhiều so với kiểu dữ liệu khác. Tốc độ thực hiện tác vụ đọc-ghi dữ liệu giữa bộ nhớ ngoài và bộ nhớ trong theo cơ chế bit được cải thiện tốt hơn so với những cơ chế khác.

Dựa trên tập kiến tạo, các thuật toán khai thác tăng cường trong Chương 2 thuộc vào nhóm kỹ thuật khai thác trực tiếp. Trong Chương 3, kỹ thuật khai thác tăng cường theo lô được phát triển. Cụ thể, tập kiến tạo của toàn cơ sở dữ liệu được tạo ra từ các tập kiến tạo thành phần, tương ứng với việc chia cơ sở dữ liệu thành các lô giao tác hoặc các lô hạng mục. Các thuật toán khai thác tăng cường theo lô giao tác hoặc lô hạng mục cũng được đề xuất.

Tiếp đó, trong Chương 4, từ các thuật toán xử lý theo lô, giải pháp song song hóa được phát triển và thực thi trên Hadoop-Spark để giải quyết bài toán với các tập dữ liệu có dung lượng lớn.

Bên cạnh đó, tập kiến tạo cũng được ứng dụng để giải quyết bài toán rút gọn tập thuộc tính đặc trưng và khai thác luồng dữ liệu.