TRANG THÔNG TIN VỀ LUẬN ÁN

Tên luận án: Khai thác tập mục hữu ích cao trên môi trường tính toán song song

Ngành: Khoa học máy tính Mã số: 9480101

Họ tên nghiên cứu sinh: Nguyễn Đắc Dzự Trình

Người hướng dẫn khoa học (ghi rõ học hàm, học vị, họ và tên):

- PGS. TS. Nguyễn Thị Thúy Loan
- TS. Phạm Thị Thiết

Cơ sở đào tạo: Trường Đại học Công nghiệp Thành phố Hồ Chí Minh

1. Tóm tắt nội dung luận án

- Tăng hiệu quả cho quá trình khai thác các tập mục hữu ích cao đóng để áp dụng và phục vụ cho bài toán tìm luật kết hợp hữu ích cao, bảo toàn lượng thông tin đã khai thác.

- Khảo sát các kỹ thuật, chiến lược kết hợp bài toán khai thác tập mục hữu ích cao với định dạng cơ sở dữ liệu phân cấp. Đề xuất mô hình song song hóa để cải thiện hiệu năng khai thác của bài toán này.

- Tối ưu hóa quá trình khai thác tập mục hữu ích cao trên định dạng cơ sở dữ liệu phân cấp với không gian tìm kiếm lớn, tận dụng hiệu quả hơn sức mạnh tính toán của các vi xử lý đa nhân phổ biến hiện nay để tăng hiệu quả của toàn bộ quá trình khai thác.

2. Những kết quả mới của luận án:

- Đề xuất các giải thuật cho phép khai thác hiệu quả các tập mục hữu ích cao sử dụng biểu diễn thu gọn và bảo toàn. Áp dụng cho định dạng cơ sở dữ liệu giao dịch truyền thống và cơ sở dữ liệu phân cấp.

- Đề xuất hai mô hình song song hóa sử dụng kiến trúc xử lý đa nhân với mức độ hiệu quả tăng dần, giúp rút ngắn thời gian thực hiện của các giải thuật khai thác tập mục hữu ích cao, tận dụng hiệu quả các vi xử lý đa nhân.

3. Các ứng dụng/khả năng ứng dụng trong thực tiễn, các vấn đề cần tiếp tục nghiên cứu:

- Giảm thiểu kích thước tập kết quả của bài toán khai thác luật kết hợp hữu ích cao trên cơ sở dữ liệu thông thường và cơ sở dữ liệu phân cấp mà vẫn bảo toàn độ chính xác.

- Mở rộng các phương pháp khai thác tập mục hữu ích cao hiện có để tận dụng tối đa sức mạnh của các vi xử lý đa nhân đang được sử dụng rộng rãi hiện nay nhằm nâng cao hiệu quả của bài toán lên đến hơn 100 lần với chi phí thấp.

- Các mô hình đề xuất cần áp dụng các chiến lược cân bằng tải hiệu quả hơn để cải thiện hiệu quả khai thác.

- Mở rộng mô hình khai thác song song đã đề xuất để hoạt động trên môi trường tính toán phân tán.

Người hướng dẫn	**Nghiên cứu sinh**
(Ký tên, họ tên)	(Ký tên, họ tên)
PGS. TS. Nguyễn Thị Thúy Loan TS. Phạm Thị Thiết	Nguyễn Đắc Dzự Trình

CONTRIBUTIONS OF THE DISSERTATION

Dissertation: Khai thác tập mục hữu ích cao trên môi trường tính toán song song

Specialization: Computer Science Code: 9480101

Ph.D. Candidate: Nguyen Dac Dzu Trinh

Supervisors:

- Assoc. Prof. Dr. Nguyen Thi Thuy Loan
- Dr. Pham Thi Thiet

Institution: Industrial University of Ho Chi Minh City

1. Summary of the dissertation content

- Enhance the efficiency of the process of mining high utility closed itemsets to apply and serve the problem of mining high utility association rules while preserving the amount of extracted information.

- Survey techniques and strategies that combine the problem of mining high utility itemsets with a hierarchical database format. Propose a parallelization model to improve the performance of this mining problem.

- Optimize the process of mining high utility itemsets on a hierarchical database format with a large search space, more effectively leveraging the computational power of modern multi-core processors to increase the efficiency of the entire mining process.

2. Contributions of the dissertation

- Propose algorithms that allow for the efficient mining of high utility itemsets using compact and lossless representations. These algorithms are applicable to both traditional transaction database formats and hierarchical database formats.

- Propose two parallelization models utilizing multi-core processing architectures with increasing levels of efficiency, helping to reduce the execution time of high utility itemset mining algorithms by up to 100 times and effectively leveraging multi-core processors.

3. Applications/practical applicability and issues that require further research:

- Reduce the result set size of the high utility association rule mining problem on both conventional databases and hierarchical databases while maintaining accuracy.

- Extend existing high utility itemset mining methods to fully leverage the power of widely used multi-core processors, aiming to enhance the efficiency of the problem-solving process at a low cost.

- Apply more effective load balancing strategies to the proposed models to improve mining efficiency further.

- Extend the proposed parallel mining model to operate in distributed computing environments.

Supervisors **Ph.D. Candidate**

Assoc. Prof. Dr. Nguyen Thi Thuy Loan Dr. Pham Thi Thiet Nguyen Dac Dzu Trinh