



THÔNG TIN VỀ LUẬN ÁN TIẾN SĨ

1. Họ và tên nghiên cứu sinh: CÙ THU THỦY
2. Giới tính: Nữ
3. Ngày sinh: 18/12/1972
4. Nơi sinh: Phú Thọ
5. Quyết định công nhận nghiên cứu sinh số: 2377/SDH, ngày 14 tháng 12 năm 2006 của Giám đốc Đại học Quốc gia Hà Nội
6. Các thay đổi trong quá trình đào tạo: Không
7. Tên đề tài luận án: Nghiên cứu phát hiện luật kết hợp hiếm và ứng dụng
8. Chuyên ngành: Hệ thống thông tin
9. Mã số: 62 48 05 01
10. Cán bộ hướng dẫn khoa học:

Cán bộ hướng dẫn 1: PGS.TS. Đỗ Văn Thành

Cán bộ hướng dẫn 2: PGS.TS. Hà Quang Thụy

11. Tóm tắt các **kết quả mới** của luận án:

Về nghiên cứu lý thuyết, luận án đề xuất phát triển một số dạng luật hiếm Sporadic, và xây dựng các thuật toán để phát hiện các dạng luật này.

Trên cơ sở dữ liệu tác vụ luận án nghiên cứu theo hướng tiếp cận đi tìm các tập hiếm đóng cho các luật hiếm thay vì việc đi tìm tất cả các tập hiếm như các nghiên cứu trước đây. Đề xuất ba thuật toán tìm các tập hiếm cho ba dạng luật hiếm là: thuật toán MCPSI phát hiện tập Sporadic tuyệt đối hai ngưỡng, MCISI phát hiện tập Sporadic không tuyệt đối hai ngưỡng và NC-CHARM phát hiện tập dữ liệu với ràng buộc mục dữ liệu âm.

Trên cơ sở dữ liệu định lượng, luận án sử dụng lý thuyết tập mờ để chuyển cơ sở dữ liệu định lượng về cơ sở dữ liệu mờ và thực hiện phát hiện luật hiếm trên cơ sở dữ liệu mờ này. Luận án đã đề xuất thuật toán MFPSI phát hiện tập Sporadic tuyệt đối hai ngưỡng mờ và MFISI phát hiện tập Sporadic không tuyệt đối hai ngưỡng mờ.

12. Khả năng ứng dụng trong thực tiễn:

Luận án đề xuất kết hợp phát hiện luật kết hợp mẫu âm và mô hình hồi quy chuyển tiếp tron trong việc xây dựng mô hình phân tích và dự báo. Luận án đã xây dựng mô hình dự báo chỉ số chứng khoán và mô hình dự báo chỉ số giá tiêu dùng Việt Nam (CPI) theo cách tiếp cận này. Những mô hình đó không chỉ hỗ trợ phân tích các quan hệ kinh tế, nhất là quan hệ nhân quả mà còn thực sự cải thiện chất lượng dự báo, độ chính xác của dự báo theo mô hình là khá cao.

13. Những hướng nghiên cứu tiếp theo:

Nghiên cứu tiếp theo của chúng tôi sẽ là:

- a) Nghiên cứu giải pháp sinh các luật hiếm có giá trị từ các tập hiếm tìm được.
- b) Áp dụng hướng phát hiện song song luật hiếm.
- c) Tiếp tục triển khai ứng dụng luật kết hợp với các phương pháp khác để phân tích và dự báo dữ liệu kinh tế.

14. Các công trình đã công bố có liên quan đến luận án:

- [1]. Cù Thu Thủy, Đỗ Văn Thành (2008), “Một giải pháp mới về phân tích thị trường chứng khoán Việt Nam”, *Tạp chí Tin học và Điều khiển học* Tập 24 (2), tr. 107-118.
- [2]. Cù Thu Thủy, Đỗ Văn Thành (2009), “Phát hiện luật kết hợp với ràng buộc mục dữ liệu âm”, *Tạp chí Tin học và Điều khiển học* Tập 25 (4), tr. 345-354.
- [3]. Đỗ Văn Thành, Phạm Thị Thu Trang, Cù Thu Thủy (2009), “Xây dựng mô hình dự báo giá bằng kết hợp mô hình hồi quy chuyển tiếp trơn và kỹ thuật phát hiện luật kết hợp”, *Kỷ yếu Hội thảo lần thứ hai trong khuôn khổ Nghị định thư Việt Nam - Thái Lan*, Đại học Kinh tế Quốc dân, tr. 308-322.
- [4]. Cù Thu Thủy, Đỗ Văn Thành (2010), “Mining Perfectly Sporadic Rules with Two Thresholds”, *In Proceedings of MASS2010*, Wuhan, China.
- [5]. Cù Thu Thủy, Đỗ Văn Thành (2010), “Mining Imperfectly Sporadic Rules with Two Thresholds”, *International Journal of Computer Theory and Engineering* Vol. 2 (5), pp. 1793-8201.
- [6]. Cù Thu Thủy, Hà Quang Thụy (2010), “Phát hiện luật kết hợp Sporadic tuyệt đối hai ngưỡng mờ”, *Kỷ yếu Hội thảo quốc gia lần thứ XIII Một số vấn đề chọn lọc của Công nghệ thông tin và Truyền thông*, Hưng Yên, tr. 263-275.
- [7]. Đỗ Văn Thành, Cù Thu Thủy, Phạm Thị Thu Trang (2010), “Building CPI Forecasting Model by Combining the Smooth Transition Regression Model and Mining Association Rules.”, *Journal on Information Technologies and Communications* Vol E-1 (7), pp.16-27.
- [8]. Cù Thu Thủy, Hà Quang Thụy (2011), “Phát hiện tập mục Sporadic không tuyệt đối hai ngưỡng mờ”, *Tạp chí Tin học và Điều khiển học* Tập 27 (2), tr. 142-153.

Ngày 08 tháng 01 năm 2013

Xác nhận của cán bộ hướng dẫn
(Kí và ghi rõ họ tên)

Ngày 08 tháng 01 năm 2013

Nghiên cứu sinh
(Kí và ghi rõ họ tên)



INFORMATION ON DOCTORAL THESIS

1. Full name: CU THU THUY
2. Sex: Female
3. Date of birth: 18/12/1972
4. Place of birth: Phu Tho
5. Admission decision number: 2377/SDH Dated: 14/12/2006 by President of Vietnam National University, Hanoi.
6. Changes in academic process: None
7. Official thesis title: Researching to mine the rare association rules and their application
8. Major: Information System
9. Code: 62 48 05 01
10. Supervisors:

Assoc. Prof. Dr. Do Van Thanh

Assoc. Prof. Dr. Ha Quang Thuy

11. Summary of the **new contributions** of the thesis:

About the theoretical research, the thesis has proposed to develop some kinds of sporadic rules and has developed algorithms to find rare itemsets for these rules.

On the transactional database, the thesis has been done on the approach to find out the closed rare itemsets for the rare rules instead of finding all of the rare itemsets like the previous researches. The thesis has been proposed three algorithms mining rare itemsets for three kinds of rare rules such as MCPSI algorithm mining closed perfectly sporadic itemsets with two thresholds, MCISI algorithm mining closed imperfectly sporadic itemsets with two thresholds and NC-CHARM algorithm mining the itemsets with negative items constraint.

For the problem of mining rare rules on quantitative database, the thesis has used the theory of fuzzy set to transfer the quantitative database into fuzzy database and to discover the rare rules relied on this fuzzy database. The thesis has proposed MFPSI algorithm to mine the fuzzy perfectly sporadic rules with two thresholds and MFISI to discover the fuzzy imperfectly sporadic rules with two thresholds.

12. Practical applicability, if any:

The thesis has proposed the combination between the mining negative association rules and smooth transition regression model on the construction of analytic and predicative model. The thesis has built a predicative model of securities index and a predicative model of Vietnam Consumption Price Index (CPI) based on this approach. Those models do not only help to analyze the economic relations, especially the cause

and effect relation, but improve highly the quality of predication, accuracy of the predication also.

13. Further research directions, if any:

Our future research will be:

- a) Finding the solutions to generate useful sporadic association rules from the sporadic itemsets.
- b) Applying the approach parallel mining rare association rules.
- c) Continue the application of the mining association rules combined with other methods to analyze and forecast economic data.

14. Thesis-related publications:

- [1]. Cu Thu Thuy, Do Van Thanh (2008), “A New Solution for Analyzing Vietnamese Stock Market”, *Journal of Computer and Cybernetics* Vol 24 (2), pp. 107-118.
- [2]. Cu Thu Thuy, Do Van Thanh (2009), “Mining Association Rules with Negative Item Constraints”, *Journal of Computer and Cybernetics* Vol 25 (4), pp. 345-354.
- [3]. Do Van Thanh, Pham Thi Thu Trang, Cu Thu Thuy (2009), “Building Price Forecasting Model by Combining the Smooth Transition Regression Model and Mining Association Rules”, *In Proceedings of the second workshop in the framework of the Protocol Vietnam – Thailand*, University of National Economy, pp. 308-322.
- [4]. Cu Thu Thuy, Do Van Thanh (2010), “Mining Perfectly Sporadic Rules with Two Thresholds”, *In Proceedings of MASS2010*, Wuhan, China.
- [5]. Cu Thu Thuy, Do Van Thanh (2010), “Mining Imperfectly Sporadic Rules with Two Thresholds”, *International Journal of Computer Theory and Engineering* Vol. 2 (5), pp. 1793-8201.
- [6]. Cu Thu Thuy, Ha Quang Thuy (2010), “Mining Fuzzy Perfectly Sporadic Association Rules with Two Thresholds”, *In Proceedings of the 13th National Conference on Information Technology and Communications*, Hung Yen, pp. 263-275.
- [7]. Do Van Thanh, Cu Thu Thuy, Pham Thi Thu Trang (2010), “Building CPI Forecasting Model by Combining the Smooth Transition Regression Model and Mining Association Rules.”, *Journal on Information Technologies and Communications* Vol E-1 (7), pp.16-27.
- [8]. Cu Thu Thuy, Ha Quang Thuy (2011), “ Mining Fuzzy Imperfectly Sporadic Itemsets with Two Thresholds”, *Journal of Computer and Cybernetics* Vol 27 (2), pp. 142-153.

Date: January 8, 2013
SUPERVISOR’S SIGNATURE

Date: January 8, 2013
RESEARCHER’S SIGNATURE