

**ĐẠI HỌC THÁI NGUYÊN  
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN VÀ TRUYỀN THÔNG**

**LÊ THỊ BÍCH THẢO**

**KHAI PHÁ LUẬT KẾT HỢP MỜ DỰA  
TRÊN ĐẠI SỐ GIA TỬ**

**LUẬN VĂN THẠC SĨ KHOA HỌC MÁY TÍNH**

**Thái Nguyên - 2013**

**ĐẠI HỌC THÁI NGUYÊN**  
**TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN VÀ TRUYỀN THÔNG**

**LÊ THỊ BÍCH THẢO**

**KHAI PHÁ LUẬT KẾT HỢP MỜ DỰA  
TRÊN ĐẠI SỐ GIA TỬ**

*Chuyên ngành: Khoa học máy tính*

*Mã số: 60.48.01*

**LUẬN VĂN THẠC SĨ KHOA HỌC MÁY TÍNH**

**NGƯỜI HƯỚNG DẪN KHOA HỌC**

**TS.Trần Thái Sơn**

**Thái Nguyên - 2012**

## LỜI CAM ĐOAN

Tên tôi là : **Lê Thị Bích Thảo**

Sinh ngày 02 tháng 7 năm 1983

Học viên cao học lớp: K9B- trường Đại học CNTT&TT Thái Nguyên

Xin cam đoan : Đề tài luận văn "***Khai phá luật kết hợp mờ dựa trên đại số gia tử***" do TS.Trần Thái Sơn hướng dẫn là công trình nghiên cứu của riêng tôi. Tất cả tài liệu tham khảo đều có nguồn gốc, xuất xứ rõ ràng.

Tôi xin cam đoan tất cả những nội dung trong luận văn đúng như nội dung trong đề cương và yêu cầu của thầy giáo hướng dẫn. Nếu sai tôi xin hoàn toàn chịu trách nhiệm trước Hội đồng khoa học và trước pháp luật.

*Thái Nguyên, tháng 01 năm 2013*

**Người cam đoan**

**Lê Thị Bích Thảo**

## LỜI CẢM ƠN

Trong quá trình làm luận văn vừa qua, dưới sự giúp đỡ và chỉ bảo nhiệt tình của TS. Trần Thái Sơn – Viện Công nghệ thông tin – Viện khoa học Việt Nam, luận văn của tôi đã được hoàn thành. Mặc dù đã cố gắng không ngừng cùng với sự tận tâm của thầy hướng dẫn nhưng do thời gian và khả năng vẫn còn hạn chế nên luận văn khó tránh khỏi những thiếu sót.

Để hoàn thành luận văn này, em xin bày tỏ lòng biết ơn sâu sắc đến TS Trần Thái Sơn – Người thầy đã tận tình giúp đỡ em trong suốt quá trình làm luận văn.

Em cũng xin bày tỏ lòng biết ơn đến ban lãnh đạo và các thầy giáo, cô giáo trong Trường Đại học Công Nghệ Thông Tin & Truyền Thông Đại Học Thái Nguyên đã giúp đỡ, tạo điều kiện tốt nhất cho em học tập và thực hiện luận văn này.

*Thái Nguyên, tháng 01 năm 2013*

**Tác giả**

**Lê Thị Bích Thảo**

## MỤC LỤC

LỜI CAM ĐOAN .....	iii
LỜI CẢM ƠN .....	iv
DANH MỤC CÁC KÝ HIỆU, CÁC CHỮ VIẾT TẮT .....	iii
DANH MỤC CÁC HÌNH ẢNH .....	iv
PHẦN MỞ ĐẦU .....	1
Chương 1: LÝ THUYẾT CHUNG VỀ TẬP MỜ VÀ LÝ THUYẾT .....	4
ĐẠI SỐ GIA TỬ .....	4
1.1. Lý thuyết chung về tập mờ.....	4
1.2. Logic mờ .....	9
1.3. Biến ngôn ngữ .....	14
1.4. Một số khái niệm cơ bản về Đại số gia tử .....	15
1.4.1. Đại số gia tử .....	17
1.4.2. Định nghĩa đại số gia tử.....	18
Chương 2: LUẬT KẾT HỢP TRONG KHAI PHÁ DỮ LIỆU .....	33
2.1. Bài toán kinh điển dẫn đến việc khai phá luật kết hợp .....	33
2.2. Khai phá luật kết hợp mờ:.....	39
Chương 3: ỨNG DỤNG ĐẠI SỐ GIA TỬ GIẢI BÀI TOÁN KHAI PHÁ DỮ LIỆU.....	41
3.1. Ứng dụng đại số gia tử trong khai phá dữ liệu. ....	41
3.1.1. Tiếp cận Đại số gia tử trong khai phá dữ liệu:.....	41
3.1.2. Thuật toán trích xuất luật kết hợp từ cơ sở dữ liệu:.....	43

3.1.3.Thuật toán giải bài toán khai phá luật kết hợp mờ dựa trên đại số gia tử .....	51
3.2 .Bài toán .....	51
3.3. Xác định đầu vào, đầu ra của bài toán .....	52
3.3.1. Thuật toán giải .....	52
3.3.2.Chương trình thử nghiệm.....	52
3.3.3. Cài đặt chương trình .....	52
3.3.4.Giao diện của chương trình.....	53
KẾT LUẬN .....	55
TÀI LIỆU THAM KHẢO.....	56
PHẦN PHỤ LỤC .....	58

## DANH MỤC CÁC KÝ HIỆU, CÁC CHỮ VIẾT TẮT

Các kí hiệu, các chữ viết tắt	Ý nghĩa
ĐSGT	Đại số gia tử
$\alpha$	Tổng độ đo tính mờ của các gia tử âm
$\beta$	Tổng độ đo tính mờ của các gia tử dương
AX, AT	Đại số gia tử
<u>AX</u>	Đại số gia tử tuyến tính đầy đủ
W	Phần tử trung hòa trong đại số gia tử

**DANH MỤC CÁC HÌNH ẢNH**

<b>Hình</b>	<b>Mô tả</b>
Hình 1	Đồ thị biểu diễn hàm thuộc của tập mờ già (old)
Hình 2	Biểu diễn bộ 2
Hình 3	Độ đo tính mờ của biến TRUTH
Hình 4	Giao diện của chương trình
Hình 5	Kết quả thực hiện chương trình thử nghiệm



## PHẦN MỞ ĐẦU

Trong những năm gần đây, việc nắm bắt được thông tin được coi là cơ sở của mọi hoạt động sản xuất, kinh doanh. Cá nhân hoặc tổ chức nào thu thập và hiểu được thông tin, và hành động dựa trên các thông tin được kết xuất từ các thông tin đã có sẽ đạt được thành công trong mọi hoạt động. Chính vì lý do đó, việc tạo ra thông tin, tổ chức lưu trữ và khai thác ngày càng trở nên quan trọng và gia tăng không ngừng.

Sự tăng trưởng vượt bậc của các cơ sở dữ liệu (CSDL) trong cuộc sống như: thương mại, quản lý và khoa học đã làm nảy sinh và thúc đẩy sự phát triển của kỹ thuật thu thập, lưu trữ, phân tích và khai phá dữ liệu... không chỉ bằng các phép toán đơn giản thông thường như: phép đếm, thống kê... mà đòi hỏi cách xử lý thông minh hơn, hiệu quả hơn. Từ đó các nhà quản lý có được thông tin có ích để tác động lại quá trình sản xuất, kinh doanh của mình... đó là tri thức. Các kỹ thuật cho phép ta khai thác được tri thức hữu dụng từ CSDL (lớn) được gọi là các kỹ thuật khai phá dữ liệu (DM – Data Mining). Khai phá luật kết hợp là một nội dung quan trọng trong khai phá dữ liệu.

Luận văn trình bày một số vấn đề về phát hiện tri thức, khai phá dữ liệu, tập trung vào vấn đề khai phá luật kết hợp và ứng dụng lý thuyết Đại số gia tử trong khai phá luật kết hợp trên CSDL.

Khai phá dữ liệu, cụ thể là trích xuất các luật kết hợp từ cơ sở dữ liệu, có xuất phát điểm từ bài toán nghiên cứu số liệu bán hàng trong siêu thị. Ở bài toán này, số liệu được biểu diễn dưới dạng bảng hai chiều, trong đó các cột thể hiện các loại mặt hàng (item), các hàng thể hiện các giao dịch (transactions) đã được tiến hành, số 1 cho thấy mặt hàng được mua, số 0 chỉ điều ngược lại. Từ bảng dữ liệu rất lớn này, người ta mong muốn rút ra được các quy luật giúp cho quản lý, kiểu như "Nếu một người đã mua bánh mì và

bơ, khả năng người đó cũng mua giảm bông là rất cao". Luật có dạng như vậy gọi là luật kết hợp và là hướng nghiên cứu quan trọng trong lĩnh vực khai phá dữ liệu. Về sau, người ta thấy sẽ là rất không đầy đủ nếu chỉ xem xét các cơ sở dữ liệu chỉ bao gồm các phân tử 0 và 1. Chẳng hạn, trong CSDL nhân sự của một cơ quan có các mục như tuổi, thu nhập.. có giá trị trong miền số thực rất rộng. Để trích xuất ra các luật kết hợp, một phương pháp thường được sử dụng là chuyển số liệu trong CSDL đã cho về CSDL chỉ chứa các giá trị 0, 1 và áp dụng các kết quả đã có. Thí dụ, trong mục "tuổi", có thể chia ra các miền "trẻ", "trung niên" và "già" với các miền giá trị tương ứng là  $[0,35]$ ,  $[36,55]$ ,  $[56,80]$  và nếu một giá trị của CSDL ban đầu rơi vào miền giá trị nào thì ta ghi 1 cho vị trí tương ứng trong CSDL chuyển đổi, ngược lại gán giá trị 0. Phương pháp này đơn giản về mặt thực thi nhưng có thể gây băn khoăn do ranh giới cứng mà người ta đưa ra khi tiến hành chuyển đổi. Chẳng hạn hai người tuổi 35 và 36 tuy rất gần nhau về mặt tuổi tác nhưng lại thuộc hai lớp khác nhau là "trẻ" và "trung niên", dẫn tới việc đưa ra những luật kết hợp có thể thiếu tính chính xác. Và người ta sử dụng cách tiếp cận mờ để khắc phục điều này, theo đó, một giá trị bất kỳ của CSDL ban đầu không chuyển đổi về giá trị 0 hoặc 1 như trên mà sẽ chuyển về một tập giá trị thực thuộc đoạn  $[0,1]$ , là độ thuộc của giá trị đã cho vào các tập mờ được xác định trước. Thí dụ, người tuổi 35 trong ví dụ trên, ở CSDL đã chuyển đổi sẽ nhận tập giá trị (trẻ, 0,8), (trung niên, 0,6), (già, 0,1). Phương pháp này, tuy dẫn tới việc xử lý phức tạp hơn nhưng dễ chấp nhận hơn về mặt trực quan và hiện đang được nhiều nhà nghiên cứu quan tâm. Mặc dù vậy, theo ý chúng tôi, phương pháp trích xuất luật kết hợp mờ vẫn có một số điểm yếu cần khắc phục. Đó là sự phụ thuộc chủ quan rất lớn vào việc lựa chọn các hàm thuộc cho các tập mờ dẫn đến việc xử lý vừa phức tạp vừa có thể thiếu chính xác. Trong bài báo này