

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

ĐẶNG THỊ THU

**XÂY DỰNG HỆ LUẬT MỜ
TỪ CƠ SỞ DỮ LIỆU - CÁCH TIẾP CẬN
THEO LÝ THUYẾT ĐẠI SỐ GIA TỬ**

LUẬN VĂN THẠC SĨ KHOA HỌC MÁY TÍNH

Chuyên ngành: KHOA HỌC MÁY TÍNH

Mã số: 60.48.01

Người hướng dẫn khoa học: TS. TRẦN THÁI SƠN

Thái Nguyên, 2014

LỜI CAM ĐOAN

Tên tôi là : **Đặng Thị Thu**

Sinh ngày 05 tháng 8 năm 1983

Học viên cao học lớp: CK11G - trường Đại học CNTT&TT
Thái Nguyên

Xin cam đoan : Đề tài luận văn ***“Xây dựng hệ luật mờ từ cơ sở dữ liệu - cách tiếp cận theo lý thuyết 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, ngày 25 tháng 8 năm 2014

Người cam đoan

Đặng Thị Thu

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, ngày 25 tháng 8 năm 2014

Tác giả

Đặng Thị Thu

MỤC LỤC

MỤC LỤC.....	i
DANH MỤC CÁC KÝ HIỆU, CÁC CHỮ VIẾT TẮT	iii
DANH MỤC CÁC HÌNH.....	iv
PHẦN MỞ ĐẦU.....	1
Chương 1: NHỮNG KIẾN THỨC CƠ BẢN VỀ LÝ THUYẾT TẬP MỜ .	3
1.1. Lý thuyết về tập mờ	3
1.1.1 Kiến thức cơ sở về tập mờ ([5])	3
1.1.2 Biến ngôn ngữ	8
1.2. Lý thuyết về Đại số gia tử ([1-3])	14
1.2.1. Những khái niệm cơ bản về đại số gia tử.....	14
1.2.2 Vấn đề định lượng ngữ nghĩa trong đại số gia tử	17
Chương 2: GIẢI THUẬT DI TRUYỀN	25
2.1. Những khái niệm cơ bản về giải thuật di truyền.....	25
2.2. Các tính chất đặc thù của thuật giải di truyền.....	28
2.3. Các bước quan trọng trong việc áp dụng giải thuật di truyền.....	29
2.4. Các phương thức biến hoá của giải thuật di truyền	29
Chương 3: XÂY DỰNG HỆ LUẬT MỜ VÀ GIẢI BÀI TOÁN HỒI QUY MỜ THEO CÁCH TIẾP CẬN CỦA ĐẠI SỐ GIA TỬ	32
3.1. Bài toán hồi quy mờ	32
3.1.1 Bài toán hồi quy mờ	32
3.1.2 Chuyển đổi CSDL số sang hệ luật mờ dựa trên lý thuyết tập mờ cổ điển	37
3.1.3 Xây dựng hệ luật mờ theo cách tiếp cận ĐSGT	41
3.2. Bài toán thiết kế tối ưu hệ luật mờ	57
3.2.1 Đặt bài toán	57
3.2.2 Tìm kiếm hệ luật tối ưu dựa trên giải thuật di truyền lai	58

3.3. Chương trình thử nghiệm.....	61
3.3.1. Cài đặt chương trình.....	61
3.3.2. Giao diện của chương trình.....	61
KẾT LUẬN CHUNG.....	73
TÀI LIỆU THAM KHẢO.....	74

DANH MỤC CÁC KÝ HIỆU, CÁC CHỮ VIẾT TẮT

Các kí hiệu, các chữ viết tắt	Ý nghĩa
ĐSGT	Đại số gia tử
A	Tổng độ đo tính mờ của các gia tử âm
B	Tổng độ đo tính mờ của các gia tử dương
AX	Đại số gia tử
\underline{AX}	Đại số gia tử tuyến tính đầy đủ
$\mu(h) fm(x)$	Độ đo tính mờ gia tử h và của hạng từ x
\tilde{Y}	Là đầu ra mờ,
\tilde{V}	Là các hệ số mờ
CSDL	Cơ sở dữ liệu
GA	Giải thuật di truyền
$\tilde{\mathcal{J}}$	Khoảng tính mờ của giá trị ngôn ngữ
X_k	Tập các hạng từ có độ dài đúng k
I_k	Hệ khoảng tính mờ mức k của các giá trị ngôn ngữ
<i>IFRG1</i>	Initial Fuzzy Rules Generation 1
<i>IFRG2</i>	Initial Fuzzy Rules Generation 2
<i>HAFRG</i>	Hedge Algebras based Fuzzy Rules Generation
<i>FPO-SGA</i>	Fuzzy Parameters Optimization - SGA
<i>RBO-SGA</i>	Rule base Optimization - SGA

DANH MỤC CÁC HÌNH

Hình	Mô tả
Hình 1	Đồ thị biểu diễn hàm thuộc của tập mờ “già” (old)
Hình 2	Độ đo tính mờ của biến TRUTH
Hình 3	Khoảng tính mờ của các hạng từ của biến TRUTH
Hình 4	Mã hóa cá thể từ không gian các lời giải của bài toán
Hình 5	Hàm định lượng dạng tam giác của các hạng từ
Hình 6:	Sơ đồ mã hóa cá thể chọn hệ luật cho thuật toán SGA

PHẦN MỞ ĐẦU

Trong cuộc sống hàng ngày hay trong công việc giảng dạy tại trường, chúng ta thường xuyên phải đưa ra những quyết định. Chẳng hạn, với một học sinh kém, ta cần có chế độ bồi dưỡng các kiến thức cơ sở mà thông thường học sinh đó bị rỗng. Với một học sinh giỏi, ta cũng cần bồi dưỡng các kiến thức, nhưng là các kiến thức mới, đòi hỏi phải tư duy tốt và tính sáng tạo trong suy nghĩ... Nói chung, cách tiến hành cụ thể là phụ thuộc vào từng học sinh và căn cứ vào kinh nghiệm giảng dạy cũng như kinh nghiệm sống của từng giáo viên và kinh nghiệm học được của đồng nghiệp, của người xung quanh.. Các kinh nghiệm này, trong tư duy của con người, có thể khái quát dưới dạng mệnh đề kiểu “ Nếu... thì..”. Thí dụ “Nếu Học lực của học sinh là Kém và Ý thức học tập của học sinh là trung bình Thì Dạy kèm theo phương án C1”; Thí dụ “Nếu Học lực của học sinh là Khá và Ý thức học tập của học sinh là Tốt Thì Dạy kèm theo phương án C2”...

Hiện tại, người ta nhận thấy, các mệnh đề dạng như trên có thể bắt gặp rất nhiều trong những lĩnh vực nghiên cứu khác nhau như điều khiển tối ưu, phân loại tự động, hồi quy. Và một hướng nghiên cứu, thuộc về khai phá dữ liệu, liên quan đến việc xây dựng các mệnh đề như vậy, mà người ta gọi là luật, để giải các bài toán khác nhau, đã và đang phát triển rất mạnh mẽ. Cụ thể, vấn đề đặt ra là từ một Cơ sở dữ liệu số (CSDL số), sử dụng các thuật toán để sinh tự động một hệ luật tối ưu (theo nghĩa gọn nhất có thể và đạt độ chính xác theo yêu cầu đặt ra). Nếu một hệ M luật được tạo ra, có dạng:

$$R_m: \text{ IF } X_i \text{ is } A_{1,j_{m,1}} \text{ AND } \dots \text{ AND } X_F \text{ is } A_{F,j_{m,F}} \text{ THEN } X_{F+1} \text{ is } A_{F+1,j_{m,F+1}}; m = 1, \dots, M,$$

trong đó X_i là các biến ngôn ngữ (như “tuổi”, “học lực”..) và $A_{i,j}$ là các giá trị biến ngôn ngữ (như “khá”, “kém”..) thì người ta gọi đó là hệ luật mờ Mamdani (Mamdani fuzzy rule-based system):

Số hóa bởi Trung tâm Học liệu – Đại học Thái Nguyên <http://www.lrc-tnu.edu.vn/>

MFRBS). MFRBS có đặc điểm khác các mô hình khác là các biến đầu vào và ra đều là mờ dưới dạng từ của ngôn ngữ tự nhiên. Đặc điểm này mang lại tính “thân thiện” với con người vì suy luận trên các từ của ngôn ngữ tự nhiên là đặc điểm của con người. Các luật cũng được biểu diễn dưới dạng quen thuộc với suy nghĩ và lập luận của con người. Ngoài ra, việc có những số liệu chính xác để xây dựng một hệ luật (không mờ) trong thời gian tính toán chấp nhận được là điều không dễ dàng. Để xây dựng MFRBS có thể có nhiều cách tiếp cận khác nhau. Trong luận văn này sử dụng cách tiếp cận của Đại số gia tử (ĐSGT), một cách tiếp cận tương đối mới và hứa hẹn cho những kết quả khả quan so với một số cách tiếp cận khác.

Được sự đồng ý của trường Đại học Công nghệ thông tin và Truyền thông với sự hướng dẫn của Thầy giáo em xin mạnh dạn nhận đề tài: **“Xây dựng hệ luật mờ từ cơ sở dữ liệu - cách tiếp cận theo lý thuyết Đại số gia tử”** làm đề tài luận văn của mình.

Luận văn có bố cục như sau:

Chương 1: Tổng quan về tập mờ và đại số gia tử

Trong chương này trình bày những kiến thức cơ bản về lý thuyết tập mờ và lý thuyết Đại số gia tử.

Chương 2: Giải thuật di truyền

Trong chương này nêu khái niệm cơ bản về giải thuật di truyền, các tính chất đặc thù của thuật giải di truyền.

Chương 3: Xây dựng hệ luật mờ và giải bài toán hồi quy mờ theo cách tiếp cận của đại số gia tử.

Trong chương này trình bày việc chuyển đổi CSDL số sang hệ luật mờ và áp dụng giải bài toán hồi quy

Chương 1

NHỮNG KIẾN THỨC CƠ BẢN VỀ LÝ THUYẾT TẬP MỜ

1.1. Lý thuyết về tập mờ

1.1.1 Kiến thức cơ sở về tập mờ ([5])

Là người đầu tiên nghiên cứu lý thuyết tập mờ, L. A. Zadeh đã có rất nhiều công trình nghiên cứu cho sự phát triển và ứng dụng. Ý tưởng nổi bật của Zadeh là từ những khái niệm trừu tượng về ngữ nghĩa của thông tin mờ, không chắc chắn như *trẻ-già*, *nhANH-chẬM*, *cao-thẤP*,... Ông đã tìm cách biểu diễn chúng bằng một khái niệm toán học, được gọi là tập mờ và được định nghĩa như sau.

Định nghĩa 1. Cho một tập vũ trụ U với các phần tử ký hiệu bởi x , $U = \{x\}$. Một tập mờ A trên U là tập được đặc trưng bởi một hàm $\mu_A(x)$ mà nó liên kết mỗi phần tử $x \in U$ với một số thực trong đoạn $[0,1]$. Giá trị hàm $\mu_A(x)$ biểu diễn mức độ thuộc của x trong A . $\mu_A(x)$ là một ánh xạ từ U vào $[0,1]$ và được gọi là hàm thuộc của tập mờ A .

Như vậy, giá trị hàm $\mu_A(x)$ càng gần tới 1 thì mức độ thuộc của x trong A càng cao. Khi A là một tập hợp kinh điển, hàm thuộc của nó, $\mu_A(x)$, chỉ nhận 2 giá trị 1 hoặc 0, tương ứng với x có nằm trong A hay không. Rõ ràng, tập mờ là sự mở rộng của khái niệm tập hợp kinh điển. Các khái niệm, phép toán trong lý thuyết tập kinh điển cũng được mở rộng cho các tập mờ.

Họ tất cả các tập mờ trên miền cơ sở U là không gian các hàm từ U vào đoạn $[0,1]$, tức là $F(U, [0,1]) = \{\mu_A : U \rightarrow [0,1]\}$, một không gian tương đối giàu về cấu trúc tính toán mà nhiều nhà nghiên cứu đã sử dụng cho việc mô phỏng các phương pháp suy luận của con người.