ĐẠI HỌC THÁI NGUYÊN

TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN VÀ TRUYỀN THÔNG

VŨ ĐỨC HẢI

PHƯƠNG PHÁP TRÍCH RÚT CÁC LUẬT MỜ PHÂN LỚP DỰA TRÊN ĐẠI SỐ GIA TỬ VÀ ỨNG DỤNG

LUẬN VĂN THẠC SĨ KHOA HỌC MÁY TÍNH

Thái Nguyên – 2015

ĐẠI HỌC THÁI NGUYÊN

TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN VÀ TRUYỀN THÔNG

VŨ ĐỨC HẢI

PHƯƠNG PHÁP TRÍCH RÚT CÁC LUẬT MỜ PHÂN LỚP DỰA TRÊN ĐẠI SỐ GIA TỬ VÀ ỨNG DỤNG

Chuyên ngành: Khoa học máy tính

Mã số: 60 48 0101

LUẬN VĂN THẠC SĨ KHOA HỌC MÁY TÍNH

NGƯỜI HƯỚNG DẪN KHOA HỌC

TS. Dương Thăng Long

Thái Nguyên – 2015

LỜI NÓI ĐẦU

Trong cuộc sống loài người, ngôn ngữ được hình thành một cách tự nhiên để giải quyết nhu cầu trao đổi thông tin với nhau. Hơn thế, nó là công cụ để con người mô tả các sự vật, hiện tượng trong thế giới thực và dựa trên đó để tư duy, lập luận đưa ra những nhận định, phán quyết nhằm phục vụ cho cuộc sống xã hội. Ngày nay khoa học và công nghệ đã có những phát triển vượt bậc, nhiều máy móc thiết bị được tạo ra đã góp phần giải phóng sức lao động của con người. Trong đó lĩnh vực công nghệ thông tin đã có những đóng góp vô cùng to lớn cho sự phát triển kinh tế - xã hội nói chung và giúp giải phóng sức lao động không chỉ là lao động chân tay mà còn cả lao động trí óc của con người nói riêng. Công nghệ thông tin đã góp phần đưa khả năng tư duy, lập luận và sự sáng tạo kiểu như bộ não người vào máy móc thiết bị để "thông minh hơn". Để thực hiện điều này, rất nhiều nhà khoa học đã và đang nghiên cứu cả về lý thuyết lẫn ứng dụng, đưa ra các phương pháp, các quy trình nhằm kế thừa, mô phỏng khả năng của con người vào các thiết bị máy móc. Trước hết, các nhà khoa học đã phải hình thức hóa toán học các vấn đề ngôn ngữ và xử lý ngôn ngữ mà con người vẫn làm. Người đi tiên phong trong lĩnh vực này là Lotfi A. Zadeh, ông đã đề xuất khái niệm mờ từ những khái niệm mơ hồ, không rõ ràng.

Cho đến nay, hệ mờ phân lớp dạng luật (FRBCS) là mô hình được nhiều tác giả quan tâm nghiên cứu và sử dụng trong khai phá dữ liệu, tìm kiếm tri thức từ dữ liệu cho bài toán phân lớp. Thế mạnh của mô hình này là có thể cung cấp được cho người dùng cuối những tri thức dạng luật dễ hiểu , dễ sử dụng đối với con người như là những tri thức của họ . Với việc sử dụng tập mờ và lôgic mờ , các nghiên cứu đều tìm kiếm phương pháp xây dựng hệ mờ phân lớp dạng luật nhằm đạt hai mục tiêu chính: thứ nhất, hiệu quả phân lớp của hệ càng cao càng tốt; thứ hai, tính phức tạp của hệ đồng thời càng nhỏ càng tốt.

Mô hình xây dựng hệ luật mờ phân lớp dựa trên đại số gia tử được đề xuất với mục tiêu xây dựng hệ luật mờ để ứng dụng phân lớp cho các mẫu dữ liệu sao cho hệ luật phải có hiệu quả phân lớp cao, càng đơn giản, dễ hiểu và tường minh đối với người dùng càng tốt.

Tên đề tài được lựa chọn là ***"Phương pháp trích rút các luật mờ phân lớp dựa trên đại số gia tử và ứng dụng"***. Nội dung của luận văn được bố cục thành các phần như sau:

Chương 1. Kiến thức cơ bản về hệ mờ và lập luận xấp xỉ.

Chương 2. Phương pháp trích rút luật mờ phân lớp dựa trên đại số gia tử.

Chương 3. Cài đặt thử nghiệm và đánh giá.

CHƯƠNG 1:
KIẾN THỨC CƠ BẢN VỀ HỆ MỜ VÀ LẬP LUẬN XẤP XỈ

1.1. Khái quát về lập luận xấp xỉ (lập luận mờ)

Từ năm 1965 Zadeh đưa ra lý thuyết tập mờ, logic mờ nhưng phải đến những thập niên cuối của thế kỷ XX lý thuyết tập mờ, logic mờ mới được đặc biệt quan tâm nghiên cứu và ứng dụng vào trong lý thuyết điều khiển, hệ thống và trí tuệ nhân tạo. Tập mờ và logic mờ dựa trên các suy luận của con người về các thông tin không đầy đủ để hiểu biết và điều khiển hệ thống. Điều khiển mờ chính là mô phỏng cách xử lý thông tin và điều khiển của con người đối với các đối tượng, do vậy điều khiển mờ đã giải quyết thành công rất nhiều vấn đề điều khiển phức tạp trước đây chưa giải quyết được.

1.1.1. Định nghĩa tập mờ

Định nghĩa 1.1: [4] Cho 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 bở 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 từU vào [0,1] và được gọi là hàm thuộc của tập mờ A[1].

Hay A được gọi là tập mờ khi và chỉ khi:

$$A = \{(x,\mu_A(x)\ x \in U,\ \mu_A(x): U \to [0,1]\} \qquad (1)$$

Trong đó $\mu_A(x)$ được gọi là hàm thuộc của tập mờ A.

Giá trị hàm $\mu_A(x)$ càng gần tới 1 thì mức độ thuộc của x trong A càng cao. Tập mờ là sự mở rộng của khái niệm tập hợp kinh điển. Khi A là tập hợp kinh điển thì A có thể được biểu diễn như sau

$$A = \{(x, \mu_A(x)\ x \in U,\ \mu_A(x): U \to \{0,1\}\} \quad (2)$$

Khi đó hàm thuộc $\mu_A(x)$ chỉ nhận hai giá trị 0 và 1.

1.1.2. Số mờ

Định nghĩa 1.2: [4] Tập mờ *A* trên đường thẳng số thực *R* là một số mờ, nếu:

1. *A* chuẩn hóa, tức là có điểm *x'* sao cho $\mu_A(x') = 1$.

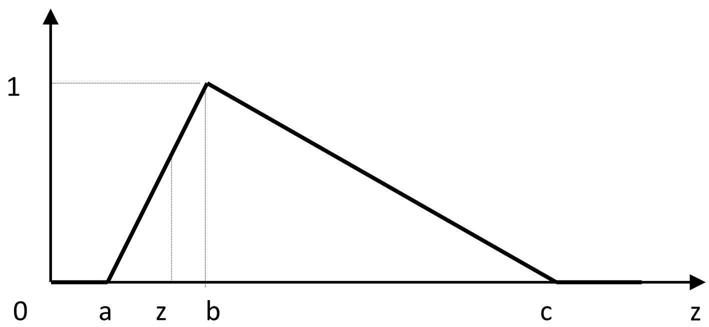
2. Ứng với mỗi $\alpha \in R$, tập mức $\{x: \mu_A(x) \geq \alpha\}$ là đoạn đóng trên *R*.

3. $\mu_A(x)$ là hàm liên tục.

Một số dạng số mờ thường được sử dụng là số mờ dạng tam giác, hình thang và dạng hàm Gauss.

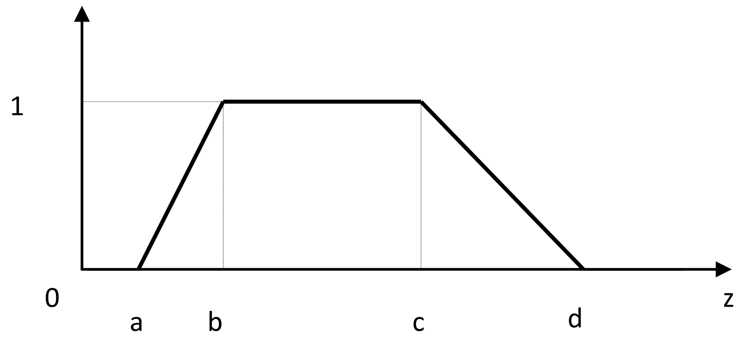
a. Số mờ dạng tam giác được xác định bởi 3 tham số. Khi đó hàm thuộc của sô mờ tam giác A(a, b, c) cho bởi:

$$\mu_A(x) = \begin{cases} 0 & \text{nếu} x \leq a \\ x - a/b - a & \text{nếu} a \leq x \leq b \\ 1 & \text{nếu} x = b \\ x - b/c - b & \text{nếu} b \leq x \leq c \\ 0 & \text{nếu} x = c \end{cases}$$



b.Số mờ hình thang A(a, b, c, d) được sác định bởi 4 tham số và hàm thuộc cho bởi:

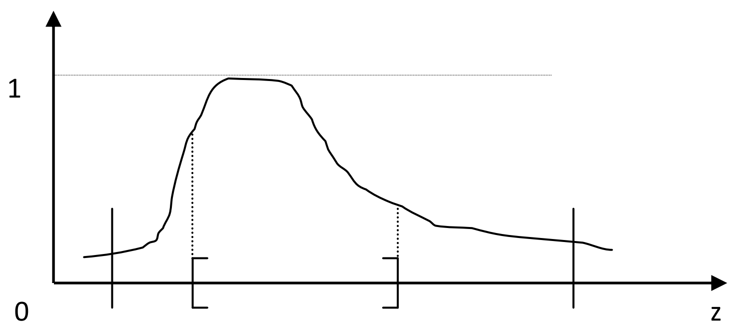
$$\mu_A(x) = \begin{cases} 0 & \text{nếu} x \leq a \\ x - a/b - a & \text{nếu} a \leq x \leq b \\ 1 & \text{nếu} b \leq x \leq c \\ x - c/d - c & \text{nếu} c \leq x \leq d \\ 0 & \text{nếu} x = d \end{cases}$$



c.Số mờ dạng hàm Gauss có hàm thuộc cho bởi:

$$\mu_A(x) = \begin{cases} e^{\frac{(x-c)^2}{(2\sigma)^2}} & \text{nếu } |\mathrm{x} - \mathrm{c}| \leq d_\alpha \\ 0 & \text{nếu } |\mathrm{x} - \mathrm{c}| \geq d_\alpha \end{cases}$$

Trong đó d_α là số dương được chọn thích hợp.



Khái niệm về phân hoạch mờ (fuzzy partition) cũng là một trong khái niệm quan trọng trong việc tiếp cận giải quyết bài toán phân lớp.

1.1.3. Định nghĩa phân hoạch mờ

Theo [4] Cho p điểm cố định $m_1<m_2<\ldots<m_p$ trong tập $U = [a, b] \subset R$. Khi đó tập gồm p tập mờ $A_1, A_2, \ldots, A_p$(với $\mu_{A_1}, \mu_{A_2}, \ldots, \mu_{A_p}$ là các hàm thuộc tương ứng) định nghĩa trên U được gọi là một phân hoạch mờ của U nếu các điều kiện sau thỏa mãn, $\forall k$=1,…,p:

1) $\mu_{A_k}(m_k) = 1$ (m_k được gọi là một điểm trong nhân của A_k);
2) Nếu $x \notin [m_{k-1}, m_{k+1}]$, $\mu_{A_k} = 0$ (trong đó $m_0 = m_1 = a$ và $m_{p+1} = m_p = b$);
3) $\mu_{A_k}(x)$ liên tục
4) $\mu_{A_k}(x)$ đơn điệu tăng trên $[m_{k-1}, m_k]$ và đơn điệu giảm trên $[m_k, m_{k+1}]$;
5) $\forall x \in U$, $\exists k$, sao cho $\mu_{A_k}(x) > 0$ (tất cả mọi điểm trong U đều thuộc một lớp của phân hoạch này với độ thuộc nào đó khác 0)

1.1.4 Các phép tính trên tập mờ Zadeh

1.1.4.1 Các phép toán tập hợp:

Cho *A*, *B* là 2 tập mờ trên cùng tập nền *U*:

Phép giao (*Intersection*):

Phép giao của tập A và B là tập mờ C được định nghĩa như sau:

$$C = A \cap B = \{(x, \mu_C(x)) | \ x \in U, \mu_C(x) = \min\{\mu_A(x), \mu_B(x)\}\}$$

Ví dụ:

Cho U = {1, 2, 3, 4, 5} và hai tập mờ A, B như sau:

A = {(1,0), (2,1), (3,0.6), (4,0.3), (5,0.2)}

B = {(1,0), (2,0.5), (3,0.7), (4,0.2), (5,0.4)}

Khi đó : C = A ∩ B = {(1,0), (2,0.5), (3,0.6), (4,0.2), (5,0.2)}

Phép hợp (*Union*):

Hợp của hai tập mờ A và B là tập mờ C được định nghĩa như sau:

$$C = A \cup B = \{\{(x, \mu_C(x))|\ x \in U,\ \ \mu_C(x) = \max\{\mu_A(x), \mu_B(x)\}\}$$

Ví dụ:

Cho U = {1, 2, 3, 4, 5} và hai tập mờ A, B như sau:

A = {(1,0), (2,1), (3,0.6), (4,0.3), (5,0.2)}

B = {(1,0), (2,0.5), (3,0.7), (4,0.2), (5,0.4)}

Khi đó : C = A ∪ B = {(1,0), (2,1), (3,0.7), (4,0.3), (5,0.4)}

Phép bù (*Complement*):

Bù của hai tập mờ A được định nghĩa như sau:

$$A^C = \{(x, \mu_{A^C}(x))\ x \in U, \mu_{A^C}(x) = 1 - \mu_A(x)\}$$

Lưu ý:

1/ $A \cup A^C \neq U$

2/ $A \cap A^C \neq 0$

3/ $(A^C)^C = A$

1.1.4.2 Phép phủ định:

Phủ định (negation) là một trong những phép toán logic cơ bản. Để suy rộng chúng ta cần tới toán tử v(Not P) xác định giá trị chân lý của Not P đối với mệnh đề P.

Định nghĩa: Hàm n: $[0, 1] \rightarrow [0, 1]$ không tăng thoả mãn các điều kiện $n(0) = 1$, $n(1) = 0$ gọi là hàm phủ định.

Hàm n là phép phủ định mạnh, nếu n giảm chặt và $n(n(x)) = x$ với mỗi x

Ví dụ: $n(x) = 1 - x$, $n(x) = 1 - x^2$

1.1.4.3 Phép hội:

Phép hội (vẫn quen gọi là phép AND – conjunction) là một trong những phép toán cơ bản nhất. Nó cũng là cơ sở để định nghĩa phép giao của hai tập mờ.

Định nghĩa 1.3: [4] Hàm $T: [0, 1] x[0, 1] \rightarrow [0, 1]$ là một phép hội hay t – chuẩn (chuẩn tam giác hay t- norm) nếu thoả mãn các điều kiện sau:

1) $T(1, x) = x$ với mọi $0 \le x \le 1$
2) T có tính giao hoán, tức là $T(x, y) = T(y, x)$ với mọi $0 \le x, y \le 1$
3) T không giảm theo nghĩa $T(x, y) \le T(u,v)$ với mọi $x \le u$, $y \le v$
4) T có tính kết hợp : $T(x, T(y, z)) = T(T(x, y), z)$ với mọi $0 \le x, y \le 1$

Ví dụ về một số t – chuẩn

$$T(x, y) = \min(x, y) \; ; \; T(x, y) = x.y \; ; \; T(x,y) = \max(x+y-1, 0)$$

1.1.4.4 Phép tuyển:

Giống như phép hội, phép tuyển hay toán tử logic OR thông thường cần thoả mãn các tính chất sau:

Định nghĩa 1.4: [4] Hàm $S : [0, 1]x[0, 1] \rightarrow [0, 1]$ gọi là phép tuyển hay là t - đối chuẩn (t – conorm) nếu thoả mãn các tiên đề sau:

1) $S(0, x) = x$ với mọi $0 \le x \le 1$
2) S có tính giao hoán: $S(x, y) = S(y, x)$ với mọi $0 \le x, y \le 1$
3) S không giảm theo nghĩa $s(x, y) \le s(u, v)$ với $x \le u$, $y \le v$