ĐẠI HỌC THÁI NGUYÊN

TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN VÀ TRUYỀN THÔNG

ĐỖ THỊ KIM DUNG

TẠO LẬP HỆ LUẬT MỜ SỬ DỤNG PHÂN CỤM TRỪ MỜ DỮ LIỆU

LUẬN VĂN THẠC SĨ KHOA HỌC MÁY TÍNH

THÁI NGUYÊN, 2017

MỤC LỤC

DANH SÁCH CÁC HÌNH VẼ iii
DANH SÁCH CÁC BẢNG BIỂU vi
MỞ ĐẦU 1
CHƯƠNG 1: CƠ SỞ LÝ THUYẾT VỀ TẬP MỜ 3
1.1 Trình bày tập Mờ. 3
1.1.1 Định nghĩa tập mờ 3
1.1.2. Một số khái niệm cơ bản của tập mờ 4
1.1.3. Biểu diễn tập mờ 5
1.2 Các phép toán trên tập mờ và hệ luật mờ 6
1.2.1 Phần bù của một tập mờ 6
1.2.4 Tích Descartes các tập mờ 8
1.2.5 Tính chất của các phép toán trên tập mờ 9
1.2.6 Hệ luật mờ 9
1.3 Lập luân xấp xỉ trong hệ mờ. 10
1.3.1 Logic mờ 10
1.3.2 Quan hệ mờ 10
1.3.3. Suy luận xấp xỉ và suy diễn mờ 11
CHƯƠNG 2. PHÂN CỤM DỮ LIỆU TRỪ MỜ. 13
2.1 Các Phương pháp phân cụm dữ liệu nói chung 13
2. 2 Phân cụm dữ liệu trừ mờ 17
2.2.1.Các thuật toán phân cụm phân hoạch 19
2.2.2 Các thuật toán phân cụm phân cấp 26
2.2.3 Các thuật toán phân cụm dựa trên mật độ. 30
2.2.4 Các thuật toán phân cụm dựa trên lưới. 32
2.2.5. Các thuật toán phân cụm dựa trên mô hình. 36
2.2.6. Các thuật toán phân cụm có dữ liệu ràng buộc 38
2.3 Các ứng dụng của phân cụm dữ liệu 39
CHƯƠNG 3. XÂY DỰNG HỆ LUẬT MỜ ỨNG DỤNG PHÂN CỤM TRỪ MỜ. 40
3.1 Xây dựng hệ luật mờ từ dữ liệu vào/ra của hệ thống 40
3.2 Ứng dụng cho bài toán lò nhiệt 45
3.2.1 Phát biểu bài toán 45
3.2.2 Mô hình động học của hệ thống lò nhiệt 47
3.3 Chương trình xử lý bài toàn và mô phỏng. 47
3.3.1 Thu thập dữ liệu vào ra của hệ thống 47
3.3.2 Hệ luật mờ cho điều khiển lò nhiệt từ phân cụm trừ 49
3.3.3 Hệ suy diễn mờ 51
3.3.4 Mô phỏng hệ thống điều khiển lò nhiệt sử dụng hệ luật mờ từ phân cụm trừ 55
KẾT LUẬN 62
TÀI LIỆU THAM KHẢO 63

DANH SÁCH CÁC HÌNH VẼ

Hình 1. 1 Hàm Thuộc có mức chuyển đổi tuyến tính. 3
Hình 1. 2 Hàm thuộc của tập B 4
Hình 1. 3 Miền xác định và miền tin cậy của tập mờ A 5

Hình 1. 4 Biểu diễn tập mờ chiều cao 6
Hình 1. 5Tập bù ***A*** của tập mờ A 6
Hình 1. 6 Hợp hai tập mờ có cùng tập nền 7
Hình 1. 7 Giao hai tập mờ có cùng tập vũ trụ 8

Hinh 2. 1 Ví dụ phân cụm của tập dữ liệu giám sát nhiệt độ lò thành 3 cụm 14
Hinh 2. 2 Các thiết lập để xác định các ranh giới các cụm ban đầu 22
Hinh 2. 3 Tính toán trọng tâm của các cụm mới 22
Hinh 2. 4 Các bước thực hiện thuật toán K- means 23
Hinh 2. 5 Thuật toán K-means chi tiết 24
Hinh 2. 6 Ví dụ về một số hình dạng cụm dữ liệu được 25
Hinh 2. 7Các chiến lược phân cụm phân cấp 27
Hinh 2. 8 Khái quát thuật toán CURE 28
Hinh 2. 9 Các cụm dữ liệu được khám phá bởi CURE 28
Hinh 2. 10 Các bước thực hiện cơ bản của thuật toán CURE 29
Hinh 2. 11 Ví dụ thực hiện phân cụm bằng thuật toán CURE 29
Hinh 2. 12 Một số hình dạng khám phá bởi phân cụm dựa trên mật độ 30
Hinh 2. 13 a) Mật độ trực tiếp, b) Đến được mật độ, c) Mật độ liên thông 31
Hinh 2. 14 Mô hình cấu trúc dữ liệu lưới 33
Hinh 2. 15 Các bước thực hiện thuật toán STING 35
Hinh 2. 16Các bước thực hiện thuật toán EM 37

Hinh 3. 1 Luật được hình thành qua phép chiếu vào không gian đầu vào *X* 40
Hinh 3. 2 Dữ liệu được phân cụm trừ , tâm cụm là điểm đơn 41
Hinh 3. 3Số lượng luật hình thành qua phan cụn trừ từ Bảng dữ liệu 3.1 45
Hinh 3. 4 Mặt suy diễn và hàm thuộc đầu vào của Bảng dữ liệu 3.1 45
Hinh 3. 5 Sơ đồ tổng quát hệ điều khiển mờ xây dựng từ dữ liệu 46
Hinh 3. 6 Bộ điều khiển mờ cho lò nhiệt ..... **Error! Bookmark not defined.**
Hinh 3. 7 Đồ thị biểu diễn số liệu thu thập được ở bảng 3.4 49
Hinh 3. 8 Hệ luật mờ hình thành sau khi phân cụm trừ 50
Hinh 3. 9 Hệ luật mờ cho điều khiển nhiệt độ 51

Hinh 3. 10 hàm liên thuộc của luật Điều khiển theo TS 52
Hinh 3. 11 Mô hình đơn giản với các hàm thuộc hình thang và tam giác cho ánh xạ vào/ ra 53
Hinh 3. 12Mô hình TS xấp xỉ từng đoạn cho hàm phi tuyến f(x).....53
Hinh 3. 13 Biểu diễn ánh xạ từ không gian vào đến không gian ra 54
Hinh 3. 14 Mặt suy diễn và các hàm thuộc đầu vào của hệ điều khiển 55
Hinh 3. 15 Đáp ứng ra (xanh) bám theo tín hiệu yêu cầu (đỏ) 61

DANH SÁCH CÁC BẢNG BIỂU

Bảng 1. 1 Bảng biểu tập mờ A 4

Bảng 3. 1 Luật mờ được xây dựng từ phân cụm trừ SC 42
Bảng 3. 2 Các cụm được xây dựng qua phân cụm trừ 43
Bảng 3. 3 Tọa độ tâm các cụm.....43
Bảng 3. 4 Dữ liệu thu thập từ đầu vào/ra của hệ thống điều khiển lò nhiệt.....48

Lời đầu tiên cho em xin kính gửi các các thầy cô bộ muôn khoa Công nghệ. Cùng toàn thể lãnh đạo thầy cô đang giảng dạy và làm việc tại trường Đại Học Công Nghệ và Truyền Thông Thái Nguyên, lời chúc sức khỏe. Em xin chúc tất cả các thầy cô giáo luôn thành công trong sự nghiệp giáo dục đào tạo cũng như mọi lĩnh vực trong cuộc sống.

Em xin chân thành Cảm ơn Thầy PGS. TS Lê Bá Dũng, người đã trực tiếp hướng dẫn và nhiệt tình chỉ bảo để em có thể hoàn thành luận văn tốt nghiệp này.

Em xin cảm ơn Ban Giám hiệu, Quý thầy cô trường Đại Đại Học Công Nghệ và Truyền Thông Thái Nguyên đã trang bị cho em một lượng kiến thức bổ ích trong quá trình tôi học tập và thực hiện đề tài. Cảm ơn bạn bè đồng nghiệp đã động viên, giúp đỡ cho em trong suốt quá trình học tập và nghiên cứu.

Cuối cùng em xin chân thành cảm ơn các thành viên trong gia đình, những người luôn dành cho tôi những tình cảm nồng ấm và chia sẻ những lúc khó khăn trong cuộc sống, luôn động viên giúp đỡ tôi trong quá trình học tập và nghiên cứu.

Do kiến thức còn hạn hẹp nên không tránh khỏi những thiếu sót trong cách hiểu, lỗi trình bày. Em rất mong nhận được sự đóng góp ý kiến của quý thầy cô và Ban lãnh đao

Em xin trân trọng cảm ơn!

DANH SÁCH CÁC CHỮ VIẾT TẮT

PCDL	Phân Cụm Dữ Liệu
KPDL	Khai Phá Dữ Liệu
CSDL	Cơ Sở Dữ Liệu

LỜI CAM ĐOAN

Em xin cam đoan đây là luận văn do em nghiên cứu và thực hiện.

Các thông số, Hình ảnh và kết quả sử dụng trong luận văn là hoàn toàn có thật và chưa từng được công bố ở bất kỳ luận văn nào khác.

Thái Nguyên, ngày 16 tháng 5 năm 2017

Tác giả luận văn:

Đỗ Thị Kim Dung

MỞ ĐẦU

Sự phát triển nhanh chóng của các hệ thống thông tin như hiện nay, thì hệ mờ được áp dụng thành công trong nhiều lĩnh vực như điều khiển tự động, phân lớp dữ liệu, phân tích việc ra quyết định, các hệ chuyên gia. Hệ luật mờ xây dựng từ tri thức nói chung hay hệ suy luận mờ nói riêng được xây dựng theo suy diễn của con người, là một phần quan trọng trong ứng dụng logic mờ cũng như trong lý thuyết tập mờ vào thực tế. Trong nhiều ứng dụng cho thiết kế các hệ thống thông minh cũng như trong xây dựng các hệ trợ giúp quyết định, hệ mờ được xây dựng theo phân lớp dữ liệu, phân cụm dữ liệu, xây dựng cây quyết định.... Hệ mờ được thực hiện từ các luật mờ, các luật mờ được xây dựng từ các tri thức của các chuyên gia trong một lĩnh vực cụ thể.

Phân cụm dữ liệu đang là một vấn đề quan tâm nghiên cứu của các tác giả trong và ngoài nước và có nhiều thuật toán phân cụm được đề xuất. Trong đó, không ít thuật toán phân cụm kết hợp với việc sử dụng giải thuật di truyền trong quá trình thực hiện. Tuy nhiên các thuật toán được đưa ra mới chỉ xét đến khía cạnh phân chia dữ liệu thành các cụm với độ chính xác cao mà chưa để tâm đến sự tối ưu các luật sử dụng.

Trong các yêu cầu đặt ra cho quá trình phân cụm thì yêu cầu về độ chính xác luôn được đặt lên hàng đầu, ngoài ra với sự kết hợp các thuật toán phân cụm và giải thuật di truyền còn thỏa mãn được tính chất tối ưu của các luật được sử dụng. Vì vậy một cách tiếp cận khác mà luận văn nêu ra đó là xây dựng hệ luật mờ cho hệ mờ từ dữ liệu là một thực tế.

Phân cụm dữ liệu là quá trình phân chia một tập dữ liệu ban đầu thành các cụm dữ liệu sao cho các phần tử trong một cụm "tương tự" (similar) với nhau và các phần tử trong các cụm khác nhau sẽ "phi tương tự" (dissimilar) với nhau. Phân cụm dữ liệu là một phương pháp học không giám sát [7][8][9].

Hiện nay, các phương pháp phân cụm đã và đang được phát triển [6] và áp dụng nhiều trong các lĩnh vực khác nhau, bao gồm: nhận dạng, phân tích dữ liệu, nghiên cứu thị trường, xử lý ảnh, [1]… Các thuật toán phân cụm cũng rất đa dạng