

**ĐẠI HỌC QUỐC GIA HÀ NỘI
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ**

ĐINH THỊ BẢO

TẬP MỜ VIỄN CẢNH VÀ ỨNG DỤNG

LUẬN VĂN THẠC SĨ CÔNG NGHỆ THÔNG TIN

Hà Nội – 2014

**ĐẠI HỌC QUỐC GIA HÀ NỘI
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ**

ĐINH THỊ BẢO

TẬP MỜ VIỄN CẢNH VÀ ỨNG DỤNG

Ngành: Công nghệ thông tin

Chuyên ngành: Hệ thống thông tin

Mã số: 60480104

LUẬN VĂN THẠC SĨ CÔNG NGHỆ THÔNG TIN

**NGƯỜI HƯỚNG DẪN KHOA HỌC: PGS.TS. NGUYỄN ĐÌNH HÓA
TS. LÊ HOÀNG SƠN**

Hà Nội – 2014

Lời cảm ơn

Trước tiên, em xin trân trọng gửi lời cảm ơn sâu sắc đến thầy giáo PGS.TS. Nguyễn Đình Hóa – Viện CNTT – ĐH Công nghệ và thầy giáo TS. Lê Hoàng Sơn – ĐH Khoa học Tự nhiên đã trực tiếp hướng dẫn và tận tình giúp đỡ em trong suốt thời gian thực hiện luận văn.

Em xin trân trọng cảm ơn các Thầy, Cô giáo trong khoa Công nghệ thông tin – Trường ĐH Công nghệ – ĐH Quốc gia Hà Nội đã tận tình chỉ dạy, cung cấp cho em những kiến thức quý báu và luôn nhiệt tình giúp đỡ, tạo điều kiện thuận lợi nhất trong suốt quá trình em học tập tại trường. Luận văn này thực hiện dưới sự tài trợ của đề tài NAFOSTED, mã số: 102.05-2014.01. Em xin trân trọng cảm ơn các Thầy, Cô giáo và các bạn thuộc Trung tâm tính toán hiệu năng cao – ĐH Khoa học Tự nhiên đã tạo điều kiện, giúp đỡ em trong suốt quá trình thực hiện luận văn này.

Đồng thời em xin chân thành cảm ơn những người thân trong gia đình cùng toàn thể bạn bè đã luôn giúp đỡ, động viên em những lúc gặp phải khó khăn trong học tập, công việc và cuộc sống.

Lời cam đoan

Những kiến thức trình bày trong luận văn là do tôi tìm hiểu, nghiên cứu và trình bày lại theo cách hiểu. Trong quá trình làm luận văn, tôi có tham khảo các tài liệu có liên quan và đã ghi rõ nguồn tài liệu tham khảo. Tôi xin cam đoan đây là công trình nghiên cứu của tôi và không sao chép của bất kỳ ai.

Hà Nội, ngày ... tháng ... năm 2014

Học viên

Đinh Thị Bảo

Mục lục

Lời cảm ơn.....	1
Lời cam đoan	2
Mục lục	3
Danh mục các ký hiệu và chữ viết tắt.....	5
Danh mục các bảng.....	6
Danh mục các hình vẽ, đồ thị	7
MỞ ĐẦU	8
CHƯƠNG 1. TỔNG QUAN VỀ TẬP MỜ.....	10
1.1. Giới thiệu	10
1.1.1. Định nghĩa tập rõ	10
1.1.2. Định nghĩa tập mờ.....	11
1.1.3. Các khái niệm liên quan.....	12
1.2. Các phép toán tập hợp trên tập mờ	12
1.2.1. Phép bằng nhau.....	12
1.2.2. Phép lấy phần bù.....	13
1.2.3. Phép chứa.....	13
1.2.4. Phép hợp.....	13
1.2.5. Phép giao.....	14
1.2.6. Một số tính chất.....	14
1.3. Một số mở rộng tập mờ.....	15
1.3.1. Tập mờ loại hai	15
1.3.2. Tập mờ trực cảm	15
1.3.3. Tập thô, tập thô mờ	16
1.4. Ứng dụng lý thuyết tập mờ	16
1.4.1. Logic mờ.....	17
1.4.2. Biến ngôn ngữ	17
1.4.3. Luật mờ.....	17
1.4.4. Phân cụm mờ.....	18
1.5. Kết luận chương.....	18
CHƯƠNG 2. TẬP MỜ VIỄN CẢNH.....	19
2.1. Giới thiệu	19
2.1.1. Định nghĩa.....	19
2.1.2. Quan hệ giữa tập mờ viễn cảnh và các tập mờ khác	19
2.1.3. Tập mờ viễn cảnh dạng khoảng	19
2.1.4. Ví dụ minh họa	20
2.2. Các phép toán trên tập mờ viễn cảnh.....	20
2.2.1. Phép chứa.....	20
2.2.2. Phép bằng nhau.....	20

2.2.3.	<i>Phép hợp</i>	20
2.2.4.	<i>Phép giao</i>	21
2.2.5.	<i>Phép lấy phần bù</i>	21
2.2.6.	<i>Tích Descartes</i>	21
2.2.7.	<i>Các phép toán trên tập mờ viễn cảnh dạng khoảng</i>	21
2.2.8.	<i>Một số tính chất</i>	22
2.3.	Khoảng cách giữa các tập mờ viễn cảnh.....	22
2.4.	Tổ hợp lồi của tập mờ viễn cảnh.....	23
2.5.	Các quan hệ mờ viễn cảnh	24
2.6.	Kết luận chương.....	25
CHƯƠNG 3. PHÂN CỤM MỜ VIỄN CẢNH		26
3.1.	Độ đo khoảng cách viễn cảnh tổng quát	26
3.1.1.	<i>Giới thiệu</i>	26
3.1.2.	<i>Định nghĩa</i>	27
3.1.3.	<i>Các định lý</i>	30
3.2.	Thuật toán phân cụm phân cấp mờ viễn cảnh.....	30
3.2.1.	<i>Định nghĩa giá trị trung bình</i>	30
3.2.2.	<i>Thuật toán HPC</i>	31
3.3.	Đánh giá	31
3.3.1.	<i>Thiết kế thực nghiệm</i>	31
3.3.2.	<i>Minh họa thuật toán</i>	35
3.3.3.	<i>So sánh chất lượng phân cụm</i>	40
3.3.4.	<i>Hệ thống HPCS</i>	46
3.4.	Kết luận chương.....	51
KẾT LUẬN		52
TÀI LIỆU THAM KHẢO		53

Danh mục các ký hiệu và chữ viết tắt

Từ viết tắt	Từ đầy đủ	Ý nghĩa
FRS	Fuzzy Rough Set	Tập thô mờ
FS	Fuzzy Sets	Tập mờ
HPC	Hierarchical Picture fuzzy Clustering	Phân cụm phân cấp mờ viễn cảnh
HPCS	Hierarchical Picture fuzzy Clustering System	Hệ thống phân cụm phân cấp mờ viễn cảnh
IFS	Intuitionistic Fuzzy Set	Tập mờ trực cảm
IHC	Intuitionistic Hierarchical fuzzy Clustering	Phân cụm phân cấp mờ trực cảm
IvPFS	Interval-valued Picture Fuzzy Sets	Tập mờ viễn cảnh dạng khoảng
PCA	Principal Components Analysis	Phân tích thành phần chính
PDM	Picture Distance Matrix	Ma trận khoảng cách viễn cảnh
PFR	Picture Fuzzy Relations	Quan hệ mờ viễn cảnh
PFS	Picture Fuzzy Sets	Tập mờ viễn cảnh hay còn gọi là Tập mờ toàn cảnh
RFS	Rough Fuzzy Set	Tập mờ thô
RS	Rough Set	Tập thô
VS	Vague Set	Tập mơ hồ

Danh mục các bảng

Bảng 3.1. Bộ dữ liệu xe hơi Quảng Châu.....	32
Bảng 3.2. Bộ dữ liệu các vật liệu xây dựng.....	33
Bảng 3.3. Một phần bộ dữ liệu bệnh tim từ UCI.....	33
Bảng 3.4. Bảng dữ liệu tâm cụm trong giai đoạn 2 thuật toán HPC2	36
Bảng 3.5. Bảng dữ liệu tâm cụm trong giai đoạn 3 thuật toán HPC2	36
Bảng 3.6. Bảng dữ liệu giá trị trung bình ở các giai đoạn của thuật toán HPC2	37
Bảng 3.7. Bảng dữ liệu giá trị trung bình ở các giai đoạn của thuật toán HPC2 sau khi áp dụng PCA.....	38
Bảng 3.8. Bảng kết quả phân cụm của IHC, CK1, CK2, HPC1, HPC2, HPC3, HPC4 trên bộ dữ liệu xe hơi Quảng Châu.....	42
Bảng 3.9. Bảng chỉ số NMI của CK1, CK2, HPC1, HPC2, HPC3, HPC4 trên bộ dữ liệu xe hơi Quảng Châu	42
Bảng 3.10. Bảng chỉ số F-Measure của CK1, CK2, HPC1, HPC2, HPC3, HPC4 trên bộ dữ liệu xe hơi Quảng Châu	42
Bảng 3.11. Bảng chỉ số Purity của CK1, CK2, HPC1, HPC2, HPC3, HPC4 trên bộ dữ liệu xe hơi Quảng Châu	42
Bảng 3.12. Bảng kết quả phân cụm của IHC, CK1, CK2, HPC1, HPC2, HPC3, HPC4 trên bộ dữ liệu vật liệu xây dựng	44
Bảng 3.13. Bảng chỉ số NMI của CK1, CK2, HPC1, HPC2, HPC3, HPC4 trên bộ dữ liệu vật liệu xây dựng	44
Bảng 3.14. Bảng chỉ số F-Measure của CK1, CK2, HPC1, HPC2, HPC3, HPC4 trên bộ dữ liệu vật liệu xây dựng	44
Bảng 3.15. Bảng chỉ số Purity của CK1, CK2, HPC1, HPC2, HPC3, HPC4 trên bộ dữ liệu vật liệu xây dựng	44
Bảng 3.16. Bảng chỉ số NMI của CK1, CK2, HPC1, HPC2, HPC3, HPC4 trên bộ dữ liệu bệnh tim	45
Bảng 3.17. Bảng chỉ số F-Measure của CK1, CK2, HPC1, HPC2, HPC3, HPC4 trên bộ dữ liệu bệnh tim.....	45
Bảng 3.18. Bảng chỉ số Purity của CK1, CK2, HPC1, HPC2, HPC3, HPC4 trên bộ dữ liệu bệnh tim	45

Danh mục các hình vẽ, đồ thị

Hình 1.1: Tập rõ và biểu diễn tập rõ.....	10
Hình 1.2: Ví dụ một tập rõ.....	10
Hình 1.3: Tập mờ và biểu diễn tập mờ.....	11
Hình 1.4: Ví dụ một tập mờ.....	11
Hình 1.5: Ví dụ hai tập mờ bằng nhau	12
Hình 1.6: Ví dụ tập mờ và phần bù tập mờ	13
Hình 1.7: Ví dụ phép chứa tập mờ	13
Hình 1.8: Ví dụ phép hợp tập mờ	14
Hình 1.9: Ví dụ phép giao tập mờ	14
Hình 3.1: Cây phân cấp của thuật toán HPC2 trên dữ liệu xe hơi Quảng Châu	37
Hình 3.2: Phân bố 2 chiều điểm dữ liệu và tâm trong giai đoạn 1 thuật toán HPC2	38
Hình 3.3: Phân bố 2 chiều điểm dữ liệu và tâm trong giai đoạn 2 thuật toán HPC2	39
Hình 3.4: Phân bố 2 chiều điểm dữ liệu và tâm trong giai đoạn 3 thuật toán HPC2	39
Hình 3.5: Phân bố 2 chiều điểm dữ liệu và tâm trong giai đoạn 4 thuật toán HPC2	40
Hình 3.6: Cây phân cấp của thuật toán CK1 trên bộ dữ liệu xe hơi Quảng Châu.....	41
Hình 3.7: Cây phân cấp của thuật toán HPC1 trên bộ dữ liệu xe hơi Quảng Châu	41
Hình 3.8: Cây phân cấp của thuật toán CK2 & HPC4 trên bộ dữ liệu vật liệu xây dựng	43
Hình 3.9: Cây phân cấp của IHC, CK1, HPC1, HPC2, HPC3 trên vật liệu xây dựng..	43
Hình 3.10: Giao diện chính của hệ thống HPCS	46
Hình 3.11: Minh họa dữ liệu đầu vào ở bảng 3.1 lưu trong tệp Excel	46
Hình 3.12: Minh họa dữ liệu đầu vào ở bảng 3.2 lưu trong tệp Excel	47
Hình 3.13: Minh họa dữ liệu đầu vào ở bảng 3.3 lưu trong tệp Excel	47
Hình 3.14: Biểu diễn dữ liệu đầu vào trên hệ thống HPCS.....	47
Hình 3.15: Các phiên bản thuật toán phân cụm trên hệ thống HPCS	48
Hình 3.16: Minh họa cây phân cấp thuật toán HPC2 trên hệ thống HPCS.....	48
Hình 3.17: Minh họa số liệu tính toán qua các giai đoạn của thuật toán HPC2 trên hệ thống HPCS	49
Hình 3.18: Minh họa phân bố điểm và tâm qua các giai đoạn của thuật toán HPC2 trên hệ thống HPCS	49
Hình 3.19: Minh họa cây phân cấp thuật toán HPC2 khi thay đổi số cụm trên hệ thống HPCS	50
Hình 3.20: Minh họa số liệu tính toán qua các giai đoạn của thuật toán HPC2 khi thay đổi số cụm trên hệ thống HPCS	50
Hình 3.21: Minh họa bảng chỉ số đánh giá trên hệ thống HPCS	51

MỞ ĐẦU

Xã hội càng phát triển thì nhu cầu con người ngày càng cao. Logic mệnh đề (hay còn gọi là logic rõ) với hai giá trị chân lý đúng/sai đã không thể giải quyết hết các bài toán phức tạp phát sinh trong thực tế. Ví dụ như quần áo như thế nào được gọi là dày hay mỏng để máy giặt biết mà có chế độ tự động sấy khô cho hợp lý? Một cách tiếp cận mới đã mang lại nhiều kết quả thực tiễn, được các nhà nghiên cứu công nghệ mới chấp nhận và tiếp tục phát triển đó là lý thuyết tập mờ, do giáo sư Lotfi A.Zadeh [28] của trường đại học California - Mỹ đưa ra đầu tiên vào năm 1965, là mở rộng tổng quát của tập rõ.

Ngày nay, ứng dụng của tập mờ và logic mờ có thể được tìm thấy trong nhiều lĩnh vực trong thực tế. Cho đến nay, lý thuyết tập mờ như các quan hệ mờ được sử dụng rộng rãi trong việc xây dựng các biến ngôn ngữ [29], các hệ hỗ trợ quyết định và phân cụm mờ. Việc sử dụng hàm thuộc cho từng phần tử trong đoạn $[0,1]$ thay vì gắn chặt một phần tử thuộc về một cụm đã giúp khắc phục việc xử lý các thông tin không đầy đủ hoặc không chính xác, nhất là trong một số ngành như Tin-sinh học, v.v. Các lý thuyết về các toán tử mờ, quan hệ mờ, các luật mờ và điều khiển mờ, v.v. đã được quan tâm bởi nhiều nhà nghiên cứu và việc ứng dụng lý thuyết này vào các bài toán thực tiễn đã được triển khai sâu rộng như: điều khiển tàu điện ngầm Hitachi tại Nhật (1987), máy giặt thông minh, máy điều hòa, robot tự động, v.v. .

Tuy nhiên, tập mờ truyền thống có một số hạn chế và vì vậy nó không thể mô hình hóa một số các sự kiện, hiện tượng trong tự nhiên. Một số mở rộng tiêu biểu của tập mờ truyền thống như: tập mờ loại 2 [29] và tập mờ trực cảm [1] đã được đề xuất sử dụng ý tưởng về mờ hóa hàm thuộc (đối với tập mờ loại 2) và thông tin do dự (đối với tập mờ trực cảm) trong định nghĩa của tập mờ đã giúp khắc phục được các nhược điểm đó. Gần đây, một loại tập mờ tổng quát của các mở rộng trên là tập mờ viễn cảnh [3] đã được Cuong & Kreinovich đề xuất nhằm hợp nhất các khái niệm này và hứa hẹn có khả năng ứng dụng lớn.

Mục tiêu của luận văn là tìm hiểu lý thuyết về tập mờ viễn cảnh (hay còn gọi là tập mờ toàn cảnh) và xây dựng một số độ đo khoảng cách viễn cảnh tổng quát được mở rộng từ độ đo của Cuong & Kreinovich. Từ đó, sử dụng các độ đo này để xây dựng thuật toán phân cụm dữ liệu mới trên tập dữ liệu PFS, cài đặt thực nghiệm nhằm kiểm chứng và đánh giá chất lượng thuật toán.

Luận văn gồm các phần sau:

Chương 1. Tổng quan về tập mờ: Chương này trình bày các khái niệm cơ bản về tập mờ nói chung và một số mở rộng của tập mờ làm cơ sở để tìm hiểu, nghiên cứu tập mờ viễn cảnh.

Chương 2. Tập mờ viễn cảnh: Tập mờ viễn cảnh PFS là mở rộng của tập mờ FS và tập mờ trực cảm IFS. Chương này trình bày định nghĩa tập mờ viễn cảnh, một số tính chất và phép toán trên tập mờ viễn cảnh PFS. Đồng thời giới thiệu khái niệm khoảng

cách giữa hai tập mờ viễn cảnh, là một trong những công cụ quan trọng trong phân cụm dữ liệu.

Chương 3. Phân cụm mờ viễn cảnh: Chương này nghiên cứu về phân cụm trên tập dữ liệu mờ viễn cảnh và đề xuất một số độ đo khoảng cách viễn cảnh tổng quát được mở rộng từ độ đo của Cuong & Kreinovich. Từ đó, sử dụng các độ đo này để xây dựng một thuật toán phân cụm dữ liệu mới trên tập dữ liệu PFS được gọi là HPC và thiết kế thực nghiệm để đánh giá chất lượng và hiệu quả của thuật toán so với thuật toán sử dụng độ đo của Cuong & Kreinovich.

CHƯƠNG 1. TỔNG QUAN VỀ TẬP MỜ

1.1. Giới thiệu

Trong lý thuyết tập hợp kinh điển (do Cantor khởi xướng), giá trị của một phần tử trong một tập hợp là 0 hoặc 1, tức là với một phần tử bất kỳ chỉ có hai khả năng là thuộc hoặc không thuộc tập hợp. Do đó, lý thuyết này không thể xử lý những dữ liệu có tính không chắc chắn, không rõ ràng.

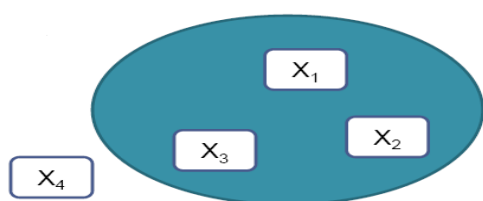
Lotfi A. Zadeh và Dieter Klaua là người sáng lập ra lý thuyết tập mờ [28], mở đầu bằng bài báo “Fuzzy Sets” trên tạp chí “Information and Control” năm 1965. Ý tưởng lý thuyết tập mờ của Zadeh là từ những khái niệm trừu tượng, không chắc chắn của thông tin như độ tuổi (trẻ – già), chiều cao (cao – thấp), nhiệt độ (nóng – lạnh), v.v. ông đã tìm ra cách biểu diễn bằng một khái niệm toán học được gọi là tập mờ FS, như là một sự khái quát của khái niệm tập hợp.

1.1.1. Định nghĩa tập rõ

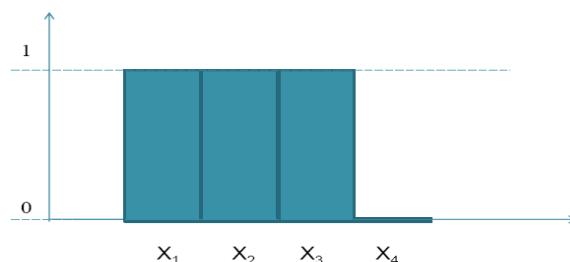
Định nghĩa 1.1 [28]: Cho tập nền X và x là phần tử của tập X . Một tập C trên tập X là một tập hợp rõ, với x là phần tử của tập hợp C , chỉ có thể có $x \in C$ hoặc $x \notin C$. Có thể sử dụng hàm $\mu(x)$ để mô tả khái niệm thuộc về. Hàm $\mu(x)$ được gọi là hàm thuộc hay hàm đặc trưng của tập hợp C .

$$\mu(x) = \begin{cases} 1 & \text{if } x \in C \\ 0 & \text{if } x \notin C \end{cases} \quad (1.1)$$

Ví dụ 1.1: $X = \{X_1, X_2, X_3, X_4\}$



$$\mu(x_1) = 1; \mu(x_2) = 1; \mu(x_3) = 1; \mu(x_4) = 0$$



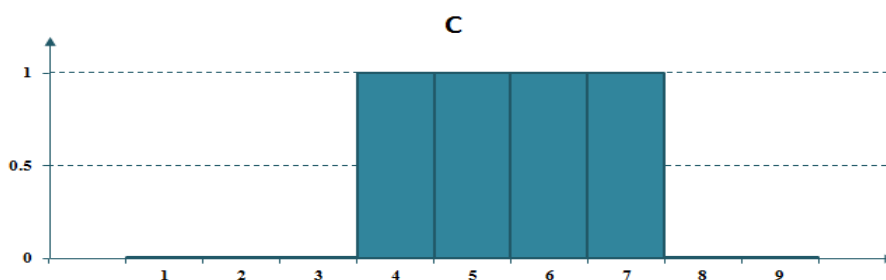
Biểu diễn tập rõ theo đồ thị

Hình 1.1: Tập rõ và biểu diễn tập rõ

Ví dụ 1.2: Cho tập $X = \{1, 2, 3, 4, 5, 6, 7, 8, 9\}$, xác định tập rõ C trên tập X với $C = \{x | 3 < x < 8\}$?

Ta có: $\mu(4) = \mu(5) = \mu(6) = \mu(7) = 1$ nên suy ra $C = \{4, 5, 6, 7\}$

Khi đó tập C được biểu diễn như Hình 1.2 sau:



Hình 1.2: Ví dụ một tập rõ

1.1.2. Định nghĩa tập mờ

Định nghĩa 1.2 [28]: Cho tập nền X và x là phần tử của tập X . Một tập mờ F trên tập X được định nghĩa bởi một hàm thành viên hay còn gọi là hàm thuộc $\mu_F(x)$ (*degree of membership*), đo “mức độ” mà phần tử x thuộc về tập F thỏa mãn điều kiện với $\forall x \in X, 0 \leq \mu_F(x) \leq 1$.

$$F = \{(x, \mu_F(x)) \mid x \in X\} \quad (1.2)$$

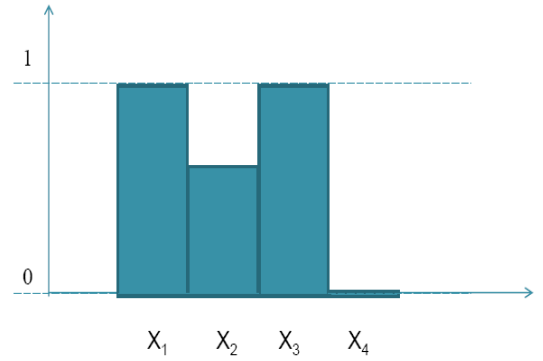
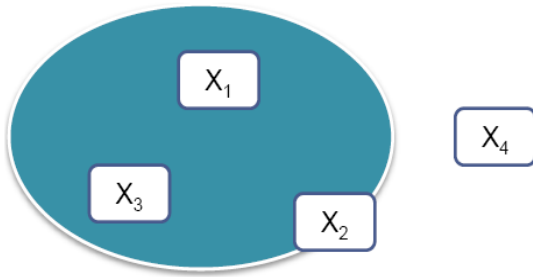
Khi $\mu_F(x) = 0$ thì $x \notin F$ hoàn toàn. Khi $\mu_F(x) = 1$ thì $x \in F$ hoàn toàn.

Tập mờ F rỗng nếu và chỉ nếu $\mu_F(x) = 0$ với $\forall x \in X$.

Tập mờ F toàn phần nếu và chỉ nếu $\mu_F(x) = 1$ với $\forall x \in X$.

Như vậy, khái niệm tập mờ là sự tổng quát hóa khái niệm tập rõ bởi hàm thuộc của nó có thể lấy giá trị bất kỳ trong khoảng $[0, 1]$, tập rõ chỉ là một tập mờ đặc biệt vì hàm thuộc $\mu_F(x)$ chỉ nhận hai giá trị 0 hoặc 1.

Ví dụ 1.3: $X = \{X_1, X_2, X_3, X_4\}$



Hàm thuộc không còn mang 2 giá trị tuyệt đối 0 hay 1, mà là giá trị thuộc đoạn $[0,1]$.

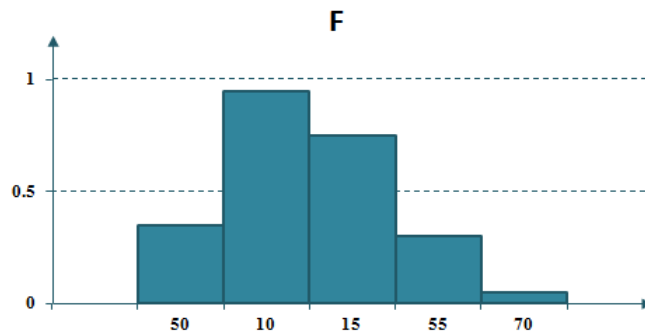
Biểu diễn tập mờ theo đồ thị.

Hình 1.3: Tập mờ và biểu diễn tập mờ

Ví dụ 1.4: Cho tập X gồm 5 người là $\{x_1, x_2, x_3, x_4, x_5\}$ tương ứng có tuổi là 50, 10, 15, 55, 70, xác định tập F là tập hợp những người “Trẻ”?

Ta có thể xây dựng hàm thuộc như sau: $\mu_F(50)=0.35, \mu_F(10)=0.95, \mu_F(15)=0.75, \mu_F(55)=0.30, \mu_F(70)=0.05$.

Khi đó tập mờ $F = \{(50, 0.35) (10, 0.95) (15, 0.75) (55, 0.30) (70, 0.05)\}$ và F được biểu diễn như Hình 1.4 sau:



Hình 1.4: Ví dụ một tập mờ

1.1.3. Các khái niệm liên quan

Định nghĩa 1.3 [28]: Độ hỗ trợ của tập mờ F , ký hiệu là $\text{supp}(F)$, bao gồm các phần tử có $\mu_F(x) > 0$, được xác định như sau:

$$\text{supp}(F) = \{x \in X \mid \mu_F(x) > 0\} \quad (1.3)$$

Định nghĩa 1.4 [28]: Lực lượng của tập mờ F , ký hiệu là $\text{card}(F)$, bao gồm tổng tất cả các $\mu_F(x)$ của các phần tử x của nó, được xác định như sau:

$$\text{card}(F) = \sum_{x \in X} \mu_F(x) \quad (1.4)$$

Ví dụ 1.5: Cho tập mờ $B = \{(3,0.2) (4,0.6) (1,0.1) (5,0.5)\}$

→ Lực lượng của tập mờ B là $\text{card}(B) = 0.2 + 0.6 + 0.1 + 0.5 = 1.4$

Định nghĩa 1.5 [28]: Quan hệ mờ R giữa hai tập X, Y là một tập mờ trên nền tích Descartes $X \times Y$, được đặc trưng bởi hàm thuộc $\mu_R(x, y) \in [0, 1]$ với $x \in X$ và $y \in Y$ và được xác định như sau:

$$R(X, Y) = \{((x, y), \mu_R(x, y)) \mid (x, y) \in X \times Y\} \quad (1.5)$$

1.2. Các phép toán tập hợp trên tập mờ

Các phép toán trên tập mờ được định nghĩa thông qua các hàm thuộc của chúng và được xây dựng tương tự như các phép toán trong tập mờ kinh điển.

Định nghĩa 1.6 [28]: Cho tập nền X và ba tập mờ A, B, C trên tập X với các hàm thuộc tương ứng là $\mu_A(x)$, $\mu_B(x)$, $\mu_C(x)$. Các phép toán tập hợp trên tập mờ được định nghĩa như sau.

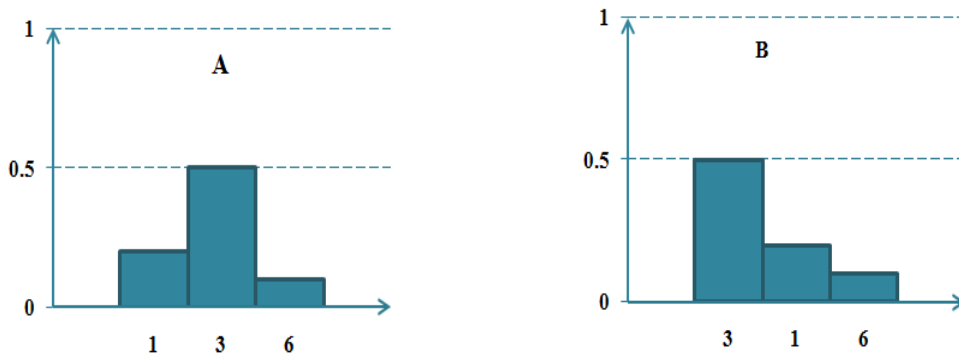
1.2.1. Phép bằng nhau

$$A = B \Leftrightarrow \mu_A(x) = \mu_B(x), \forall x \in X \quad (1.6)$$

Từ công thức (1.6) ta thấy hai tập mờ là bằng nhau, khi và chỉ khi mọi phần tử của tập này cũng thuộc tập kia và có cùng hàm thuộc. Điều này hoàn toàn tương tự khái niệm bằng nhau của hai tập hợp cổ điển.

Ví dụ 1.6: Cho tập mờ $A = \{(1,0.2) (3,0.5) (6,0.1)\}$ và $B = \{(3,0.5) (1,0.2) (6,0.1)\}$

→ $A = B$ vì $\mu_A(1) = \mu_B(1) = 0.2$ và $\mu_A(3) = \mu_B(3) = 0.5$ và $\mu_A(6) = \mu_B(6) = 0.1$



Hình 1.5: Ví dụ hai tập mờ bằng nhau

1.2.2. Phép lấy phần bù

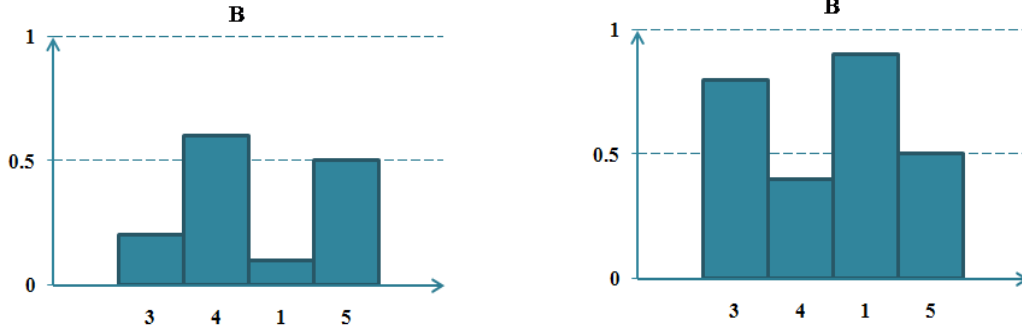
Bù của một tập mờ thể hiện mức độ một phần tử không thuộc tập đó là bao nhiêu. Nói cách khác, một phần tử càng ít có khả năng thuộc vào tập mờ A thì càng có nhiều khả năng thuộc vào phần bù của A.

Đối với tập con mờ A trên X, phần bù của A, ký hiệu là \bar{A} được xác định như sau:

$$C = \bar{A} \Leftrightarrow \mu_C(x) = 1 - \mu_A(x), \forall x \in X \quad (1.7)$$

Ví dụ 1.7: Cho tập mờ B = {(3,0.2) (4,0.6) (1,0.1) (5,0.5)}

$$\rightarrow \bar{B} = \{(3, 0.8) (4, 0.4) (1, 0.9) (5, 0.5)\}$$



Hình 1.6: Ví dụ tập mờ và phần bù tập mờ

1.2.3. Phép chứa

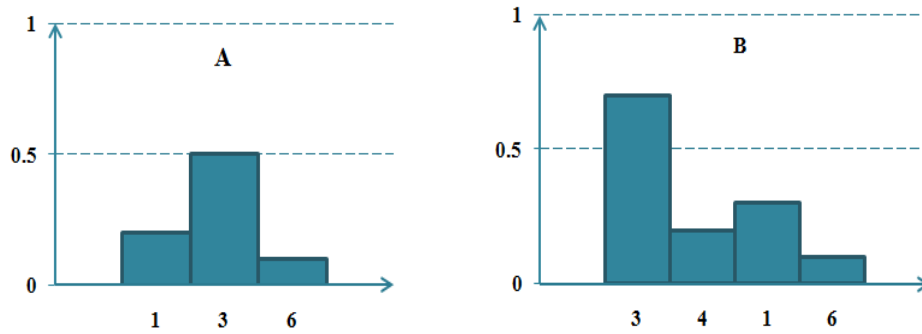
$$A \subseteq B \Leftrightarrow \mu_A(x) \leq \mu_B(x), \forall x \in X \quad (1.8)$$

Từ công thức (1.8) ta thấy tập mờ B chứa tập mờ A nếu một phần tử bất kỳ thuộc A thì cũng thuộc B với độ thuộc trên B không thấp hơn độ thuộc của phần tử đó trên A, điều này cũng tương tự như đối với các tập hợp cổ điển.

Ví dụ 1.8: Cho tập mờ A = {(1,0.2) (3,0.5) (6,0.1)}

$$\text{và } B = \{(3,0.7) (4,0.2) (1,0.3) (6,0.1)\}$$

$$\rightarrow A \subseteq B \text{ vì } \mu_A(1) < \mu_B(1), \mu_A(3) < \mu_B(3) \text{ và } \mu_A(6) = \mu_B(6)$$



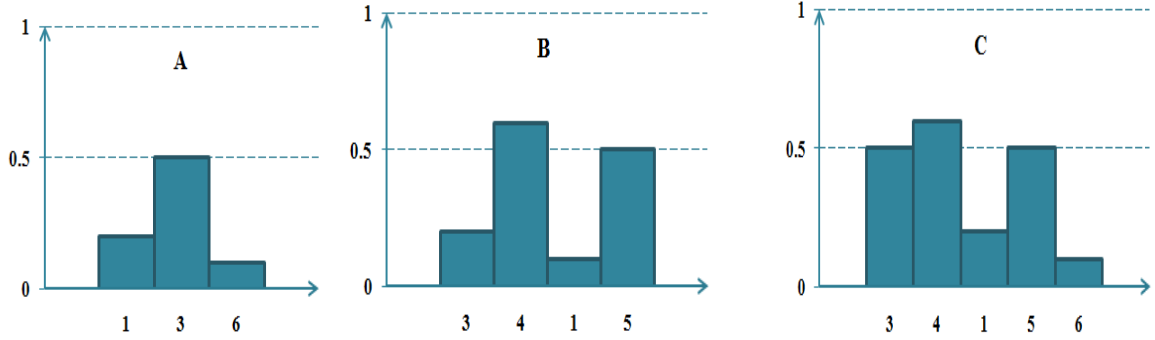
Hình 1.7: Ví dụ phép chứa tập mờ

1.2.4. Phép hợp

Hợp của hai tập mờ ($A \cup B$) thể hiện mức độ một phần tử thuộc về một trong hai tập mờ là bao nhiêu, được xác định như sau:

$$\begin{aligned} C = A \cup B &\Leftrightarrow \mu_C(x) = \max[\mu_A(x), \mu_B(x)], \forall x \in X \\ &\Leftrightarrow \mu_C(x) = \mu_A(x) \vee \mu_B(x) \end{aligned} \quad (1.9)$$

Ví dụ 1.9: Cho tập mờ $A = \{(1,0.2) (3,0.5) (6,0.1)\}$
 và $B = \{(3,0.2) (4,0.6) (1,0.1) (5,0.5)\}$
 $\rightarrow A \cup B = \{(3,0.5) (4,0.6) (1,0.2) (5,0.5) (6,0.1)\}$



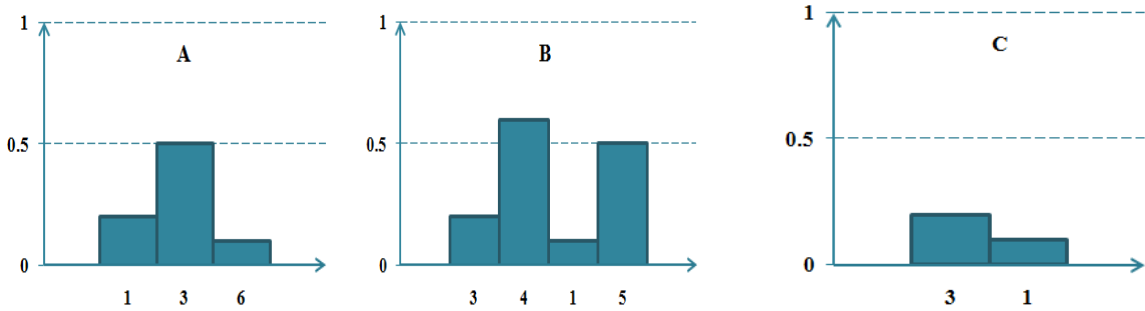
Hình 1.8: Ví dụ phép hợp tập mờ

1.2.5. Phép giao

Giao của hai tập mờ $(A \cap B)$ thể hiện mức độ một phần tử thuộc về cả hai tập mờ là bao nhiêu, được xác định như sau:

$$\begin{aligned} C = A \cap B &\Leftrightarrow \mu_C(x) = \min[\mu_A(x), \mu_B(x)], \forall x \in X \\ &\Leftrightarrow \mu_C(x) = \mu_A(x) \wedge \mu_B(x) \end{aligned} \quad (1.10)$$

Ví dụ 1.10: Cho tập mờ $A = \{(1,0.2) (3,0.5) (6,0.1)\}$
 và $B = \{(3,0.2) (4,0.6) (1,0.1) (5,0.5)\}$
 $\rightarrow A \cap B = \{(3,0.2) (1,0.1)\}$



Hình 1.9: Ví dụ phép giao tập mờ

1.2.6. Một số tính chất

- Một số công thức của tập hợp cổ điển chỉ liên quan đến phép hợp và phép giao, thì cũng đúng đối với các tập mờ, chẳng hạn như: $A \cap B \subseteq A \subseteq A \cup B$, $A \cap X = A$, $A \cup X = X$, v.v.
- Một số tính chất giao hoán, kết hợp và phân bố đối với phép hợp, phép giao của tập mờ tương tự giống như trong lý thuyết tập hợp cổ điển như: $A \cap B = B \cap A$, $A \cup B = B \cup A$, $(A \cap B) \cap C = A \cap (B \cap C)$, $(A \cup B) \cup C = A \cup (B \cup C)$, $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$, $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$, v.v.
- Một số tính chất về phần bù của tập mờ:

Đối với tập rõ thì $A \cap \bar{A} = \emptyset$ và $A \cup \bar{A} = X$, nhưng đối với tập mờ thì tính chất này không hẳn đúng. Nếu \bar{A} là phần bù của tập mờ A trên X thì $A \cap \bar{A} \neq \emptyset$ và $A \cup \bar{A} \neq X$. Đây là điểm khác nhau quan trọng giữa các tập rõ và tập mờ.

Ngoài ra, có một số tính chất khác về phần bù của tập cổ điển vẫn đúng cho các tập mờ như: $\overline{\bar{A}} = A$, $\overline{A \cup B} = \bar{A} \cap \bar{B}$, $\overline{A \cap B} = \bar{A} \cup \bar{B}$, v.v.

1.3. Một số mở rộng tập mờ

Lý thuyết tập mờ thông thường (tập mờ loại một) gặp phải vấn đề đó là để phát triển một hệ logic mờ phải xây dựng hàm thuộc cho các tập mờ sử dụng trong hệ, hay phải mô tả sự không chắc chắn bằng các hàm thuộc rõ ràng và chính xác, với độ thuộc là một số thực trên đoạn $[0, 1]$.

Năm 1975, Zadeh đã giới thiệu khái niệm tập mờ loại hai [29] nhằm giải quyết vấn đề gặp phải của tập mờ thông thường. Với tập mờ loại hai thì độ thuộc là một tập mờ loại một trên đoạn $[0, 1]$, nó cho phép biểu diễn độ thuộc bằng các giá trị mờ, các giá trị ngôn ngữ chứ không phải các giá trị số chính xác.

Atanassov đưa ra khái niệm tập mờ trực cảm IFS [1]. Gau và Buehrer (1993) đưa ra khái niệm tập mờ hồ VS và đã được chứng minh rằng VS có ý nghĩa tương tự IFS.

Ngoài ra, Pawlak (1982) đề xuất khái niệm tập thô RS, sau đó bằng cách phối hợp tập mờ và tập thô, Nanra và Majumdar (1992) đã đưa ra khái niệm tập thô mờ FRS, còn Banerjee và Pal (1996) đưa ra khái niệm tập mờ thô RFS.

1.3.1. Tập mờ loại hai

Định nghĩa 1.7 [28]: Một tập mờ loại hai trên tập nền X , ký hiệu là \tilde{F} , được định nghĩa bởi hàm thuộc $\mu_{\tilde{F}}(x, u)$ thỏa mãn điều kiện với $\forall x \in X, u \in J_x \subseteq [0, 1]$.

$$\tilde{F} = \{((x, u), \mu_{\tilde{F}}(x, u)) \mid \forall x \in X, \forall u \in J_x \subseteq [0, 1]\} \quad (1.11)$$

Tập mờ loại hai thường được sử dụng trong trường hợp khó xác định chính xác độ thuộc của các phần tử trong không gian.

1.3.2. Tập mờ trực cảm

Định nghĩa 1.8 [1]: Một tập mờ trực cảm trên tập nền X , ký hiệu là A , được định nghĩa bởi hàm thuộc $\mu_A(x) \in [0, 1]$ là mức độ khẳng định độ thuộc của x trong A (*degree of membership/validity*) và độ không thuộc $\nu_A(x) \in [0, 1]$ là mức độ phủ định độ thuộc của x trong A (*degree of non-membership/non-validity*) thỏa mãn điều kiện với $\forall x \in X, 0 \leq \mu_A(x) + \nu_A(x) \leq 1$.

$$A = \{(x, \mu_A(x), \nu_A(x)) \mid x \in X\} \quad (1.12)$$

Với $x \in X$, $\pi_A(x) = 1 - \mu_A(x) - \nu_A(x)$ được gọi là độ không chắc chắn của x trong A (*degree of indeterminacy/uncertainty*), là thành phần đại diện cho lượng thông tin bị thiếu trong việc xác định khả năng x thuộc vào A .

Trường hợp với $\forall x \in X, \pi_A(x) = 0$ thì $\nu_A(x) = 1 - \mu_A(x)$, khi đó tập mờ trực cảm IFS có dạng $\{(x, \mu_A(x), 1 - \mu_A(x)) \mid \forall x \in X\}$, và thường được biểu diễn ở dạng $\{(x, \mu_A(x)) \mid \forall x \in X\}$. Vậy với $\forall x \in X, \pi_A(x) = 0$ thì IFS trở thành tập mờ FS.

Phần bù của A, ký hiệu A^c , được xác định như sau:

$$A^c = \{(x, \nu_A(x), 1 - \mu_A(x)) \mid x \in X\} \quad (1.13)$$

Định nghĩa 1.9 [1]: Cho hai tập không rỗng X, Y. Một quan hệ mờ trực cảm, ký hiệu là R, là một tập mờ trực cảm trên $X \times Y$, được định nghĩa như sau:

$$R = \{((x, y), \mu_R(x, y), \nu_R(x, y)) \mid x \in X, y \in Y\} \quad (1.14)$$

Với $\mu_R(x, y) \in [0, 1], \nu_R(x, y) \in [0, 1]$ thỏa mãn điều kiện:

$$\forall (x, y) \in (X \times Y), 0 \leq \mu_R(x, y) + \nu_R(x, y) \leq 1 \quad (1.15)$$

1.3.3. Tập thô, tập thô mờ

Định nghĩa 1.10: Cho R là quan hệ tương đương trên tập nền không rỗng X. Cặp giá trị $W = (X, R)$ được gọi là một không gian xấp xỉ. Với tập con A của X, dựa trên quan hệ R xây dựng các xấp xỉ dưới $\underline{R}(A)$ và xấp xỉ trên $\overline{R}(A)$ của tập con A trên W như sau:

$$\underline{R}(A) = \{x \in X \mid R(x) \subseteq A\} \quad (1.16)$$

$$\overline{R}(A) = \{x \in X \mid R(x) \cap A \neq \emptyset\} \quad (1.17)$$

Cặp $(\underline{R}(A), \overline{R}(A))$ được gọi là tập thô RS, hay A xác định thô theo quan hệ R nếu $\underline{R}(A) \neq \overline{R}(A)$. Tập thô RS xác định $\Leftrightarrow \underline{R}(A) \neq \emptyset$ và $\overline{R}(A) \neq X$.

Ví dụ 1.11: Cho tập nền $X = \{1, 2, 3, \dots, 10\}$, $x_1 R x_2$ (x_1 quan hệ tương đương R với x_2) nếu x_1, x_2 có cùng số dư khi chia cho 3 và một tập con của X là $A = \{2, 3, 6, 8, 9\}$.

Ta có các quan hệ R như sau: $[1] = \{1, 4, 7, 10\}$; $[2] = \{2, 5, 8\}$; $[3] = \{3, 6, 9\}$.

Khi đó, xấp xỉ dưới của A là $\underline{R}(A) = \{3, 6, 9\}$ và xấp xỉ trên của A là $\overline{R}(A) = \{2, 3, 5, 6, 8, 9\}$.

Như vậy A xác định thô theo quan hệ R hay cặp $(\underline{R}(A), \overline{R}(A))$ là một tập thô.

Định nghĩa 1.11: Cho S là tập tất cả các tập thô, cặp $(\underline{R}(B), \overline{R}(B)) \in S$. Một tập thô mờ B trên tập X được mô tả bởi cặp ánh xạ $\mu_{\underline{R}(B)} \in [0, 1]$, $\mu_{\overline{R}(B)} \in [0, 1]$ và $\mu_{\underline{R}(B)} \leq \mu_{\overline{R}(B)}$ với $\forall x \in B$. Khi đó một tập thô mờ FRS B có thể được viết như sau:

$$B = \{(x, \mu_{\underline{R}(B)}(x), \mu_{\overline{R}(B)}(x)) \mid x \in B\} \quad (1.18)$$

Phần bù của B, ký hiệu là B^c được xác định như sau:

$$B^c = \{(x, 1 - \mu_{\overline{R}(B)}(x), 1 - \mu_{\underline{R}(B)}(x)) \mid x \in B\} \quad (1.19)$$

1.4. Ứng dụng lý thuyết tập mờ

Ứng dụng đầu tiên của lý thuyết tập mờ hay logic mờ là vào công nghiệp được thực hiện tại trường Queen Mary, London – Anh vào khoảng sau năm 1970. Ebrahim Mamdani đã dùng logic mờ để điều khiển một máy hơi nước mà trước đây ông ấy

không thể thực hiện được bằng các kỹ thuật cổ điển. Tại Đức, Hans Zimmermann đã dùng logic mờ cho các hệ quyết định. Kể từ năm 1980, logic mờ đạt được nhiều thành công trong các ứng dụng ra quyết định và phân tích dữ liệu ở Châu Âu.

Vào những năm 1980, các công ty Nhật đã bắt đầu dùng logic mờ vào kỹ thuật điều khiển. Một trong những ứng dụng đầu tiên dùng logic mờ là nhà máy xử lý nước của Fuji Electric vào năm 1983 và hệ thống tàu điện ngầm Hitachi vào năm 1987. Sau đó, logic mờ được dùng nhiều trong các ứng dụng thuộc lĩnh vực điều khiển thông minh, xử lý dữ liệu (như máy ảnh, máy quay phim, xe hơi Mishubishi, v.v.), điều khiển tự động hóa (như máy giặt, máy điều hòa, v.v.) hay được dùng để tối ưu nhiều quá trình hóa học, sinh học, v.v. Các sản phẩm thương mại dùng logic mờ ngày càng phổ biến.

Từ những thành công đạt được, logic mờ đã trở thành một kỹ thuật thiết kế “chuẩn” và được áp dụng rộng rãi. Ngày nay, ứng dụng của lý thuyết tập mờ có thể được tìm thấy trong nhiều lĩnh vực trong thực tế.

1.4.1. Logic mờ

Logic mờ được phát triển từ lý thuyết tập mờ để thực hiện lập luận xấp xỉ thay vì lập luận chính xác theo logic vị từ cổ điển. Logic mờ cho phép hàm thuộc có giá trị trong khoảng $[0,1]$ và ở hình thức ngôn từ, khái niệm không chính xác như “hơi hơi”, “gần như”, “khá là”, “rất”, v.v.

Ví dụ 1.12: “Cô ấy sống rất hạnh phúc”

“Chiếc máy này chạy khá là tốt”

1.4.2. Biến ngôn ngữ

Biến ngôn ngữ (Linguistic Variables) là thành phần chủ đạo trong các hệ thống dùng logic mờ. Các luật trong hệ logic mờ mô tả tri thức trong hệ và chúng dùng các biến ngôn ngữ như là từ vựng để mô tả các tầng điều khiển của hệ. Biến ngôn ngữ là biến nhận các giá trị ngôn ngữ (Linguistic Terms) chẳng hạn như “già”, “trẻ” và “trung niên”, v.v. trong đó, mỗi giá trị ngôn ngữ thực chất là một tập mờ xác định bởi một hàm thuộc và khoảng giá trị số tương ứng, chẳng hạn giá trị ngôn ngữ “trung niên” là một tập mờ có hàm thuộc dạng hình tam giác cân xác định trong khoảng độ tuổi $[25,55]$. Các tập này có thể xếp phủ lên nhau, chẳng hạn, một người ở tuổi 50 có thể thuộc cả tập mờ “trung niên” lẫn tập mờ “già”, với mức độ thuộc với mỗi tập là khác nhau.

1.4.3. Luật mờ

Một luật mờ là một biểu thức IF – THEN được phát biểu ở dạng ngôn ngữ tự nhiên thể hiện sự phụ thuộc nhân quả giữa các biến.

Cho 3 tập mờ A, B, C và 3 biến ngôn ngữ (Linguistic Variables) x, y, z. Luật mờ có dạng:

- IF x is A THEN y is B.
- IF x is A AND y is B THEN z is C

- IF x is A OR y is B THEN z is C

Ví dụ 1.12: **IF** nhiệt độ là lạnh **AND** giá dầu là rẻ **THEN** sương ẩm nhiều.

Trong đó: “nhiệt độ”, “giá dầu” và “sương ẩm” là các biến x,y,z. Còn “lạnh”, “rẻ”, “nhiều” là các giá trị hay chính là tập mờ A, B, C.

1.4.4. Phân cụm mờ

Bài toán phân cụm mờ được ứng dụng rất nhiều như trong việc nhận dạng mẫu (vân tay, ảnh), xử lý ảnh, y học (phân loại bệnh lý, triệu chứng), v.v. Việc sử dụng logic để phân cụm dữ liệu mềm dẻo hơn rất nhiều so với phân cụm rõ. Nó cho phép một đối tượng có thể thuộc vào một hay nhiều cụm khác nhau, được biểu diễn thông qua khái niệm hàm thuộc hay độ thuộc.

1.5. Kết luận chương

Chương một trình bày cơ bản về tập mờ nói chung và một số phép toán tập hợp trên tập mờ. Sau đó, giới thiệu sơ lược một số mở rộng của tập mờ và ứng dụng của tập mờ trong thực tế.

CHƯƠNG 2. TẬP MỜ VIỄN CẢNH

2.1. Giới thiệu

Chương 1 đã giới thiệu về lý thuyết tập mờ FS và một số mở rộng của nó. Trong [3] giới thiệu một khái niệm mới đó là tập mờ viễn cảnh PFS, đây là mở rộng trực tiếp của tập mờ FS và tập mờ trực cảm IFS của Antanassov. Chương 2 dưới đây sẽ đi vào tìm hiểu cụ thể tập mờ viễn cảnh PFS.

2.1.1. Định nghĩa

Định nghĩa 2.1 [3]: Một tập mờ viễn cảnh A trên tập X là đối tượng được định nghĩa như sau:

$$A = \{(x, \mu_A(x), \eta_A(x), \nu_A(x)) \mid x \in X\} \quad (2.1)$$

Trong đó, $\mu_A(x) \in [0,1]$ là mức độ khẳng định độ thuộc (*degree of positive membership*) của x trong A, $\eta_A(x) \in [0,1]$ là mức độ trung lập (*degree of neutral membership*) của x trong A và $\nu_A(x) \in [0,1]$ là mức độ phủ định độ thuộc (*degree of negative membership*) của x trong A, thỏa mãn điều kiện sau:

$$(\forall x \in X) \quad (\mu_A(x) + \eta_A(x) + \nu_A(x) \leq 1) \quad (2.2)$$

Với $x \in X$, $\xi_A(x) = 1 - (\mu_A(x) + \eta_A(x) + \nu_A(x))$ được gọi là mức độ “từ chối” (*degree of refusal membership*) của x trong A.

Trường hợp với $\forall x \in X$, $\xi_A(x) = 1$ và $\mu_A(x) = \eta_A(x) = \nu_A(x) = 0$ thì tập mờ A không có phần tử nào, hay tập mờ A rỗng.

2.1.2. Quan hệ giữa tập mờ viễn cảnh và các tập mờ khác

Tập mờ viễn cảnh PFS chính là tổng quát hóa của tập mờ FS và tập mờ trực cảm IFS.

Trường hợp với $\forall x \in X$, $\xi_A(x) = 0$ thì $\eta_A(x) = 1 - (\mu_A(x) + \nu_A(x))$, khi đó tập mờ viễn cảnh PFS sẽ có dạng $\{(x, \mu_A(x), 1 - (\mu_A(x) + \nu_A(x)), \nu_A(x)) \mid \forall x \in X\}$, và thường được biểu diễn dưới dạng $\{(x, \mu_A(x), \nu_A(x)) \mid \forall x \in X\}$. Vậy với $\forall x \in X$, $\xi_A(x) = 0$ thì PFS trở thành IFS.

Trường hợp với $\forall x \in X$, $\xi_A(x) = 0$ và $\eta_A(x) = 1 - (\mu_A(x) + \nu_A(x)) = 0$ thì PFS trở thành FS.

2.1.3. Tập mờ viễn cảnh dạng khoảng

Giả sử ký hiệu $\text{int}([0,1])$ là tập của tất cả các khoảng con đóng trong $[0,1]$.

Định nghĩa 2.2 [3]: Cho $[a_1, b_1], [a_2, b_2] \in \text{int}([0,1])$. Ta định nghĩa:

$$[a_1, b_1] \leq [a_2, b_2] \Leftrightarrow a_1 \leq a_2, b_1 \leq b_2 \quad (2.3)$$

$$[a_1, b_1] \tilde{\pi} [a_2, b_2] \Leftrightarrow a_1 \leq a_2, b_1 \geq b_2 \quad (2.4)$$

$$[a_1, b_1] = [a_2, b_2] \Leftrightarrow a_1 = a_2, b_1 = b_2 \quad (2.5)$$

$$[a_1, b_1] \vee [a_2, b_2] \Leftrightarrow [a_1 \vee a_2, b_1 \vee b_2] \quad (2.6)$$

$$[a_1, b_1] \wedge [a_2, b_2] \Leftrightarrow [a_1 \wedge a_2, b_1 \wedge b_2] \quad (2.7)$$

$$[a_1, b_1] \times [a_2, b_2] \Leftrightarrow [a_1 \times a_2, b_1 \times b_2] \quad (2.8)$$

Với \vee, \wedge tương ứng là ký hiệu của phép toán max, min.

Với $\alpha \geq 0$, ta có $\alpha[a_1, b_1] = [\alpha a_1, \alpha b_1]$. Nếu $0 \leq \alpha \leq 1$, thì $\alpha[a_1, b_1] \in \text{int}([0,1])$.

Định nghĩa 2.3 [3]: Một tập mờ viễn cảnh dạng khoảng A trên tập X, ký hiệu là IvPFS, là một đối tượng được định nghĩa như sau:

$$A = \{(x, M_A(x), L_A(x), N_A(x)) \mid x \in X\} \quad (2.9)$$

Với $M_A(x) = [(M_{AL}(x)), M_{AU}(x)] \in \text{int}([0,1])$,

$L_A(x) = [(L_{AL}(x)), L_{AU}(x)] \in \text{int}([0,1])$,

$N_A(x) = [(N_{AL}(x)), N_{AU}(x)] \in \text{int}([0,1])$

thỏa mãn điều kiện sau:

$$(\forall x \in X) (\sup M_A(x) + \sup L_A(x) + \sup N_A(x) \leq 1) \quad (2.10)$$

2.1.4. Ví dụ minh họa

Các mô hình dựa trên tập mờ viễn cảnh (PFS) có thể phù hợp trong các tình huống khi chúng ta phải đối mặt với các ý kiến liên quan đến các câu hỏi dạng: có (yes), tránh (abstain), không (no), từ chối (refusal).

Trong một cuộc Bầu cử, Ban bầu cử phát 500 phiếu bầu cho các ứng cử viên. Kết quả bầu cử có thể chia thành bốn nhóm: “vote for” (đồng ý) có 300 phiếu, “abstain” (bỏ phiếu trống) có 64 phiếu, “vote against” (phản đối) có 115 phiếu và “refusal of voting” (phiếu bị từ chối) có 21 phiếu. Nhóm “abstain” để lá phiếu trắng nghĩa là không đồng ý cũng không phản đối, nhưng vẫn tham gia bỏ phiếu. Nhóm “refusal of voting” bao gồm các lá phiếu không hợp lệ hoặc các lá phiếu không tham gia bầu cử. Đây là một ví dụ xảy ra trong thực tế và IFS không thể đáp ứng được vì thành viên trung lập (nhóm “abstain”) không tồn tại.

2.2. Các phép toán trên tập mờ viễn cảnh

Định nghĩa 2.4 [3]: Cho tập nền X và hai tập mờ viễn cảnh A và B trên tập X. Các phép toán trên tập mờ viễn cảnh được định nghĩa như sau:

2.2.1. Phép chứa

$$A \subseteq B \Leftrightarrow \mu_A(x) \leq \mu_B(x) \text{ and } \eta_A(x) \leq \eta_B(x) \text{ and } \upsilon_A(x) \geq \upsilon_B(x), \forall x \in X \quad (2.11)$$

Ví dụ 2.1: Cho hai tập mờ viễn cảnh: $A = \{(1, 0.3, 0.1, 0.5) (3, 0.4, 0.2, 0.3)\}$

và $B = \{(1, 0.8, 0.4, 0) (3, 0.4, 0.4, 0.1) (6, 0.5, 0.3, 0)\}$

$\rightarrow A \subseteq B$ vì $\mu_A(1) < \mu_B(1)$ và $\eta_A(1) < \eta_B(1)$ và $\upsilon_A(1) > \upsilon_B(1)$, $\mu_A(3) < \mu_B(3)$ và $\eta_A(3) < \eta_B(3)$ và $\upsilon_A(3) > \upsilon_B(3)$

2.2.2. Phép bằng nhau

$$A = B \Leftrightarrow A \subseteq B \text{ and } B \subseteq A \quad (2.12)$$

2.2.3. Phép hợp

$$A \cup B = \{(x, \max(\mu_A(x), \mu_B(x)), \min(\eta_A(x), \eta_B(x)), \min(\upsilon_A(x), \upsilon_B(x))) \mid \forall x \in X\} \quad (2.13)$$

Ví dụ 2.2: Cho hai tập mờ viễn cảnh: $A = \{(1,0.3,0.1,0.5) (3,0.4,0.2,0.3)\}$
và $B = \{(1,0.8,0.4,0) (3,0.4,0.4,0.1) (6,0.5,0.3,0)\}$
 $\rightarrow A \cup B = \{(1,0.8,0.1,0)(3,0.4,0.2,0.1)(6,0.5,0,0)\}$

2.2.4. Phép giao

$$A \cap B = \{(x, \min(\mu_A(x), \mu_B(x)), \min(\eta_A(x), \eta_B(x)), \max(v_A(x), v_B(x))) \mid \forall x \in X\} \quad (2.14)$$

Ví dụ 2.3: Cho hai tập mờ viễn cảnh: $A = \{(1,0.3,0.1,0.5) (3,0.4,0.2,0.3)\}$
và $B = \{(1,0.8,0.4,0) (3,0.4,0.4,0.1) (6,0.5,0.3,0)\}$
 $\rightarrow A \cap B = \{(1,0.3,0.1,0.5)(3,0.4,0.2,0.3)\}$

2.2.5. Phép lấy phần bù

$$\text{co}A = \bar{A} = \{(x, v_A(x), \eta_A(x), \mu_A(x)) \mid \forall x \in X\} \quad (2.15)$$

Ví dụ 2.4: Cho tập mờ viễn cảnh: $A = \{(1,0.3,0.1,0.5) (3,0.4,0.2,0.3)\}$
 $\rightarrow \bar{A} = \{(1,0.5,0.1,0.3)(3,0.3,0.2,0.4)\}$

2.2.6. Tích Descartes

Cho hai tập X_1 và X_2 và A, B là hai tập mờ viễn cảnh tương ứng trên X_1 và X_2

$$A = \{(x, \mu_A(x), \eta_A(x), v_A(x)) \mid \forall x \in X_1\}$$

$$\text{và } B = \{(y, \mu_B(y), \eta_B(y), v_B(y)) \mid \forall y \in X_2\}$$

Tích Descartes của hai tập mờ viễn cảnh A, B được định nghĩa như sau:

$$A \times_1 B = \{(x, y), \mu_A(x) \cdot \mu_B(y), \eta_A(x) \cdot \eta_B(y), v_A(x) \cdot v_B(y)) \mid \forall x \in X_1, y \in X_2\} \quad (2.16)$$

$$A \times_2 B = \{(x, y), \mu_A(x) \wedge \mu_B(y), \eta_A(x) \wedge \eta_B(y), v_A(x) \vee v_B(y)) \mid \forall x \in X_1, y \in X_2\} \quad (2.17)$$

2.2.7. Các phép toán trên tập mờ viễn cảnh dạng khoảng

Định nghĩa 2.5 [3]: Cho tập nền X và hai tập mờ viễn cảnh dạng khoảng IvPFS trên tập X là A và B . Với \vee, \wedge tương ứng là ký hiệu của phép toán max, min. Các phép toán trên tập mờ viễn cảnh dạng khoảng được định nghĩa như sau:

– Phép chứa

$$A \subseteq B \Leftrightarrow M_A(x) \leq M_B(x) \text{ and } L_A(x) \leq L_B(x) \text{ and } N_A(x) \geq N_B(x), \forall x \in X \quad (2.18)$$

– Phép bằng nhau

$$A = B \Leftrightarrow A \subseteq B \text{ and } B \subseteq A \quad (2.19)$$

– Phép hợp

$$A \cup B = \{(x, M_A(x) \vee M_B(x), L_A(x) \vee L_B(x), N_A(x) \vee N_B(x)) \mid \forall x \in X\} \quad (2.20)$$

– Phép giao

$$A \cap B = \{(x, M_A(x) \wedge M_B(x), L_A(x) \wedge L_B(x), N_A(x) \wedge N_B(x)) \mid \forall x \in X\} \quad (2.21)$$

– Phép lấy phần bù

$$\text{co}A = \bar{A} = \{(x, N_A(x), L_A(x), M_A(x)) \mid \forall x \in X\} \quad (2.22)$$

– Tích Descartes

Cho A là một tập mờ viễn cảnh dạng khoảng trên tập X_1 và B là một tập mờ viễn cảnh dạng khoảng trên tập X_2 . Tích Descartes của hai tập mờ viễn cảnh dạng khoảng A, B được định nghĩa như sau:

$$A \times_1 B = \{(x, y), M_A(x) \times M_B(y), L_A(x) \times L_B(y), N_A(x) \times N_B(y) \mid \forall x \in X_1, y \in X_2\} \quad (2.23)$$

$$A \times_2 B = \{(x, y), M_A(x) \wedge M_B(y), L_A(x) \wedge L_B(y), N_A(x) \vee N_B(y) \mid \forall x \in X_1, y \in X_2\} \quad (2.24)$$

2.2.8. Một số tính chất

Mệnh đề 1: Cho tập mờ viễn cảnh A, B, C.

$$(a) \quad \text{If } A \subseteq B \text{ and } B \subseteq C \text{ then } A \subseteq C \quad (2.25)$$

$$(b) \quad \overline{\overline{A}} = A \quad (2.26)$$

(c) Các phép toán \cap và \cup có tính chất giao hoán, kết hợp và phân phối

(d) Các phép toán \cap , phần bù (Co) và \cup thỏa mãn luật De Morgan

Mệnh đề 2: Cho ba tập X_1, X_2, X_3 và bốn tập mờ viễn cảnh PFS (hoặc tập mờ viễn cảnh dạng khoảng IvPFS) A, B (trên X_1), C (trên X_2) và D (trên X_3)

$$(a) \quad A \times_1 C = C \times_1 A \quad (2.27)$$

$$(b) \quad (A \times_1 C) \times_1 D = A \times_1 (C \times_1 D) \quad (2.28)$$

$$(c) \quad (A \cup B) \times_1 C = (A \times_1 C) \cup (B \times_1 C) \quad (2.29)$$

$$(d) \quad (A \cap B) \times_1 C = (A \times_1 C) \cap (B \times_1 C) \quad (2.30)$$

Mệnh đề 3: Với tích Descartes của hai tập mờ viễn cảnh như sau:

$$A \times_2 B = \{(x, y), \mu_A(x) \wedge \mu_B(y), \eta_A(x) \wedge \eta_B(y), \nu_A(x) \vee \nu_B(y) \mid \forall x \in X_1, y \in X_2\} \quad (2.31)$$

Tương tự, ta có:

$$(a) \quad A \times_2 C = C \times_2 A \quad (2.32)$$

$$(b) \quad (A \times_2 C) \times_2 D = A \times_2 (C \times_2 D) \quad (2.33)$$

$$(c) \quad (A \cup B) \times_2 C = (A \times_2 C) \cup (B \times_2 C) \quad (2.34)$$

$$(d) \quad (A \cap B) \times_2 C = (A \times_2 C) \cap (B \times_2 C) \quad (2.35)$$

Mệnh đề 4: Cho ba tập X_1, X_2, X_3 và bốn tập mờ viễn cảnh dạng khoảng (IvPFS) A, B (trên X_1), C (trên X_2) và D (trên X_3). Với tích Descartes của hai tập mờ viễn cảnh dạng khoảng như sau:

$$A \times_2 B = \{(x, y), M_A(x) \wedge M_B(y), L_A(x) \wedge L_B(y), N_A(x) \vee N_B(y) \mid \forall x \in X_1, y \in X_2\} \quad (2.36)$$

Tương tự, ta có:

$$(a) \quad A \times_2 C = C \times_2 A \quad (2.37)$$

$$(b) \quad (A \times_2 C) \times_2 D = A \times_2 (C \times_2 D) \quad (2.38)$$

$$(c) \quad (A \cup B) \times_2 C = (A \times_2 C) \cup (B \times_2 C) \quad (2.39)$$

$$(d) \quad (A \cap B) \times_2 C = (A \times_2 C) \cap (B \times_2 C) \quad (2.40)$$

2.3. Khoảng cách giữa các tập mờ viễn cảnh

Định nghĩa 2.6 [3]: Khoảng cách giữa hai tập mờ viễn cảnh A, B trên tập nền $X = \{x_i\}$ với $i = \overline{1, N}$ là:

– Khoảng cách Hamming chuẩn hóa, ký hiệu là $d_p(A, B)$, xác định như sau:

$$d_p(A, B) = \frac{1}{N} \sum_{i=1}^N (|\mu_A(x_i) - \mu_B(x_i)| + |\eta_A(x_i) - \eta_B(x_i)| + |\nu_A(x_i) - \nu_B(x_i)|) \quad (2.41)$$

– Khoảng cách Euclidean chuẩn hóa, ký hiệu là $e_p(A,B)$, xác định như sau:

$$e_p(A,B) = \sqrt{\frac{1}{N} \sum_{i=1}^N ((\mu_A(x_i) - \mu_B(x_i))^2 + (\eta_A(x_i) - \eta_B(x_i))^2 + (v_A(x_i) - v_B(x_i))^2)} \quad (2.42)$$

Ví dụ 2.5: Cho ba tập mờ viễn cảnh sau:

$$A = \{(0.8, 0.1, 0)/x_1, (0.4, 0.2, 0.3)/x_2, (0.5, 0.3, 0)/x_3\}$$

$$B = \{(0.3, 0.3, 0.2)/x_1, (0.7, 0.1, 0.1)/x_2, (0.4, 0.3, 0.2)/x_3\}$$

$$C = \{(0.3, 0.4, 0.1)/x_1, (0.6, 0.2, 0.1)/x_2, (0.4, 0.3, 0.1)/x_3\}$$

$$\rightarrow d_p(A,B) = 1/3 * [(|0.8-0.3|+|0.4-0.7|+|0.5-0.4|) + (|0.1-0.3|+|0.2-0.1|+|0.3-0.3|) + (|0-0.2| + |0.3-0.1|+|0-0.2|)] = 0.6$$

$$\text{Tương tự: } d_p(B,C) = 0.5/3, d_p(A,C) = 0.5$$

$$\rightarrow e_p(A,B)$$

$$= \sqrt{\frac{1}{3} \{[(0.8-0.3)^2 + (0.4-0.7)^2 + (0.5-0.4)^2] + [(0.1-0.3)^2 + (0.2-0.1)^2 + (0.3-0.3)^2] + [(0-0.2)^2 + (0.3-0.1)^2 + (0-0.2)^2]\}}$$

$$= \sqrt{\frac{0.52}{3}}$$

$$\text{Tương tự, } e_p(B,C) = \sqrt{\frac{0.05}{3}}, e_p(A,C) = \sqrt{0.15}$$

2.4. Tổ hợp lỗi của tập mờ viễn cảnh

Tổ hợp lỗi của tập mờ viễn cảnh là một phép toán quan trọng trong toán học, là công cụ hữu ích trong phân tích lỗi, không gian tuyến tính và tối ưu hóa lỗi. Trong phần này, tổ hợp lỗi được định nghĩa với một số mệnh đề đơn giản như sau.

Định nghĩa 2.7 [3]: Cho hai tập mờ viễn cảnh (PFS) A, B trên tập X và số thực θ với $0 \leq \theta \leq 1$. Với mỗi giá trị θ , tổ hợp lỗi của tập mờ viễn cảnh được định nghĩa như sau:

$$C_\theta(A,B) = \{(x, \mu_{C_\theta}(x), \eta_{C_\theta}(x), v_{C_\theta}(x)) \mid x \in X\} \quad (2.43)$$

$$\text{với } \forall x \in X, \mu_{C_\theta}(x) = \theta \cdot \mu_{A_\theta}(x) + (1-\theta) \cdot \mu_{B_\theta}(x),$$

$$\forall x \in X, \eta_{C_\theta}(x) = \theta \cdot \eta_{A_\theta}(x) + (1-\theta) \cdot \eta_{B_\theta}(x),$$

$$\forall x \in X, v_{C_\theta}(x) = \theta \cdot v_{A_\theta}(x) + (1-\theta) \cdot v_{B_\theta}(x).$$

Tương tự, ta định nghĩa tổ hợp lỗi của tập mờ viễn cảnh dạng khoảng IvPFSs.

Định nghĩa 2.8 [3]: Cho hai tập mờ viễn cảnh dạng khoảng (IvPFS) A, B trên tập X và số thực θ với $0 \leq \theta \leq 1$. Với mỗi giá trị θ , tổ hợp lỗi của tập mờ viễn cảnh dạng khoảng được định nghĩa như sau:

$$C_\theta(A,B) = \{(x, M_{C_\theta}(x), L_{C_\theta}(x), N_{C_\theta}(x)) \mid x \in X\} \quad (2.44)$$

với

$$\forall x \in X, M_{C_\theta}(x) = [(M_{C_{\theta L}}(x) = \theta \cdot M_{A_{\theta L}}(x) + (1-\theta) \cdot M_{B_{\theta L}}(x)), M_{C_{\theta U}}(x) = \theta \cdot M_{A_{\theta U}}(x) + (1-\theta) \cdot M_{B_{\theta U}}(x))]$$

$$\forall x \in X, L_{C_\theta}(x) = [(L_{C_{\theta L}}(x) = \theta \cdot L_{A_{\theta L}}(x) + (1-\theta) \cdot L_{B_{\theta L}}(x)), L_{C_{\theta U}}(x) = \theta \cdot L_{A_{\theta U}}(x) + (1-\theta) \cdot L_{B_{\theta U}}(x))]$$

$$\forall x \in X, N_{C_\theta}(x) = [(N_{C_{\theta L}}(x) = \theta \cdot N_{A_{\theta L}}(x) + (1-\theta) \cdot N_{B_{\theta L}}(x)), N_{C_{\theta U}}(x) = \theta \cdot N_{A_{\theta U}}(x) + (1-\theta) \cdot N_{B_{\theta U}}(x))]$$

Mệnh đề 5: Cho hai tập mờ viễn cảnh A, B (hoặc hai tập mờ viễn cảnh dạng khoảng IvPFS) và số thực θ với $0 \leq \theta \leq 1$. Ta có:

$$\circ \text{ If } \theta = 1 \text{ then } C_\theta(A, B) = A \text{ and If } \theta = 0 \text{ then } C_\theta(A, B) = B \quad (2.45)$$

$$\circ \text{ If } A \subseteq B \text{ then } \forall \theta, A \subseteq C_\theta(A, B) \subseteq B \quad (2.46)$$

$$\circ \text{ If } A \supseteq B \text{ and } \theta_1 \geq \theta_2 \text{ then } C_{\theta_1}(A, B) \supseteq C_{\theta_2}(A, B) \quad (2.47)$$

Mệnh đề 6: Cho ba tập mờ viễn cảnh A, B, D (hoặc ba tập mờ viễn cảnh dạng khoảng IvPFS) và số thực θ với $0 \leq \theta \leq 1$. Ta có:

$$\circ C_\theta(A \cap B, D) = C_\theta(A, D) \cap C_\theta(B, D) \quad (2.48)$$

$$\circ C_\theta(A \cup B, D) = C_\theta(A, D) \cup C_\theta(B, D) \quad (2.49)$$

2.5. Các quan hệ mờ viễn cảnh

Quan hệ mờ là một trong những khái niệm quan trọng trong lý thuyết tập mờ và lý thuyết hệ mờ, cần thiết để phát triển một cách tiếp cận mới cho các vấn đề như ra quyết định đầu tư theo từng thời kỳ, chuẩn đoán y tế, đánh giá động hiệu quả hệ thống quân sự, v.v. Tiếp theo sẽ giới thiệu một số kết quả sơ bộ trên quan hệ mờ viễn cảnh.

Định nghĩa 2.9 [3]: Cho X, Y, Z là các tập không rỗng. Một quan hệ mờ viễn cảnh là một tập con mờ viễn cảnh trên $X \times Y$, ký hiệu là R , được xác định như sau:

$$R = \{((x, y), \mu_R(x, y), \eta_R(x, y), \upsilon_R(x, y)) \mid x \in X, y \in Y\} \quad (2.50)$$

Với $\mu_R(x, y) \in [0, 1]$, $\eta_R(x, y) \in [0, 1]$, $\upsilon_R(x, y) \in [0, 1]$ thỏa mãn điều kiện với $\forall (x, y) \in (X \times Y)$, $\mu_R(x, y) + \eta_R(x, y) + \upsilon_R(x, y) \leq 1$.

Định nghĩa 2.10 [3]: Cho $PFR(X \times Y)$ là tập tất cả các tập con mờ viễn cảnh trên $X \times Y$ và $R \in PFR(X \times Y)$, quan hệ nghịch đảo R^{-1} giữa Y và X được định nghĩa như sau:

$$\mu_{R^{-1}}(y, x) = \mu_R(x, y), \eta_{R^{-1}}(y, x) = \eta_R(x, y), \upsilon_{R^{-1}}(y, x) = \upsilon_R(x, y), \forall (x, y) \in (X \times Y) \quad (2.51)$$

Định nghĩa 2.11 [3]: Cho R và P là hai quan hệ mờ viễn cảnh giữa X và Y , với $\forall (x, y) \in (X \times Y)$, ta có định nghĩa:

$$R \leq P \Leftrightarrow (\mu_R(x, y) \leq \mu_P(x, y)) \& (\eta_R(x, y) \leq \eta_P(x, y)) \& (\upsilon_R(x, y) \geq \upsilon_P(x, y)) \quad (2.52)$$

$$R \vee P = \{((x, y), (\mu_R(x, y) \vee \mu_P(x, y)), (\eta_R(x, y) \wedge \eta_P(x, y)), (\upsilon_R(x, y) \wedge \upsilon_P(x, y))) \mid x \in X, y \in Y\} \quad (2.53)$$

$$R \wedge P = \{((x, y), (\mu_R(x, y) \wedge \mu_P(x, y)), (\eta_R(x, y) \wedge \eta_P(x, y)), (\upsilon_R(x, y) \vee \upsilon_P(x, y))) \mid x \in X, y \in Y\} \quad (2.54)$$

$$R_c = \{((x, y), \upsilon_R(x, y), \eta_R(x, y), \mu_R(x, y)) \mid x \in X, y \in Y\} \quad (2.55)$$

Mệnh đề 7: Cho $R, P, Q \in PFR(X \times Y)$, ta có:

$$(a) \quad (R^{-1})^{-1} = R \quad (2.56)$$

$$(b) \quad R \leq P \Rightarrow R^{-1} \leq P^{-1} \quad (2.57)$$

$$(c1) \quad (R \vee P)^{-1} = R^{-1} \vee P^{-1} \quad (2.58)$$

$$(c2) \quad (R \wedge P)^{-1} = R^{-1} \wedge P^{-1} \quad (2.59)$$

$$(d1) \quad R \wedge (P \vee Q) = (R \wedge P) \vee (R \wedge Q) \quad (2.60)$$

$$(d2) \quad R \wedge (P \vee Q) = (R \wedge P) \vee (R \wedge Q) \quad (2.61)$$

$$(e) \quad R \wedge P \leq R, R \wedge P \leq P \quad (2.62)$$

$$(f1) \quad \text{if } (R \geq P) \& (R \geq Q) \text{ then } R \geq P \vee Q \quad (2.63)$$

$$(f2) \quad \text{if } (R \leq P) \& (R \leq Q) \text{ then } R \leq P \wedge Q \quad (2.64)$$

Ngoài ra, nội dung kết hợp các quan hệ mờ viễn cảnh và các chứng minh có thể tham khảo trong tài liệu [3].

2.6. Kết luận chương

Chương 2 đã giới thiệu khái niệm tập mờ viễn cảnh PFS, một số phép toán trên tập mờ viễn cảnh và một số tính chất. Tiếp theo đó đề cập tới khái niệm khoảng cách giữa hai tập mờ viễn cảnh, tổ hợp lồi của tập mờ viễn cảnh. Đồng thời, các quan hệ mờ viễn cảnh cũng đã được đưa ra.

CHƯƠNG 3. PHÂN CỤM MỜ VIỄN CẢNH

Chương 2 đã giới thiệu một số độ đo khoảng cách giữa các tập mờ viễn cảnh trong các công thức (2.41 – 2.42). Độ đo khoảng cách là một trong những công cụ quan trọng nhất trong phân cụm để xác định mức độ quan hệ giữa hai đối tượng và ảnh hưởng tới chất lượng phân cụm. Việc lựa chọn các độ đo khoảng cách ảnh hưởng tới chất lượng phân cụm của thuật toán. Vì vậy, ý tưởng chính trong chương này là phát triển một số độ đo khoảng cách viễn cảnh tổng quát và sử dụng các độ đo khoảng cách này để đề xuất một phương pháp phân cụm phân cấp mờ viễn cảnh mới gọi là HPC thực hiện phân cụm trên tập dữ liệu mờ viễn cảnh PFS. Những nội dung này sẽ được trình bày chi tiết dưới đây.

3.1. Độ đo khoảng cách viễn cảnh tổng quát

3.1.1. Giới thiệu

Độ đo khoảng cách trong một số tập mờ, đặc biệt là FS và IFS, đã được nghiên cứu rộng rãi trong nhiều tài liệu.

Diamond & Kloeden [4] đã đưa ra các độ đo khoảng cách giữa hai tập mờ A, B trên tập nền $X = \{x_i\}$ với $i = \overline{1, N}$ đã được sử dụng rộng rãi gồm:

- Độ đo khoảng cách Hamming ($p=1$) và Euclidean ($p=2$)

$$d(A, B) = \left(\sum_{i=1}^N |\mu_A(x_i) - \mu_B(x_i)|^p \right)^{1/p} \quad (3.1)$$

- Độ đo khoảng cách Hamming chuẩn hóa ($p=1$) và Euclidean chuẩn hóa ($p=2$)

$$d_n(A, B) = \left(\frac{1}{N} \sum_{i=1}^N |\mu_A(x_i) - \mu_B(x_i)|^p \right)^{1/p} \quad (3.2)$$

- Độ đo Hausdorff

$$d_h(A, B) = \max \{ |\mu_A(x_i) - \mu_B(x_i)| \} \quad (3.3)$$

Với $\mu_A(x_i), \mu_B(x_i) \in [0, 1]$ ($i = \overline{1, N}$) tương ứng là các hàm thuộc của A và B.

Papakostas et al. [13] and Xu & Xia [24] đã tổng quát hóa các độ đo khoảng cách mới nhất cho tập IFS, tổng kết các nghiên cứu của Atanassov [1], Chen [2], Dengfeng & Chuntian [5], Grzegorzewski [6], Hung & Yang [8-9], Hatzimichailidis et al. [7], Liang & Shi [10], Li et al. [11], Mitchell [12], Szmidt & Kacprzyk [15-17], Wang & Xin [19], Xu et al. [20-26] and Yang & Chiclana [27]. Atanassov [1] đã kết hợp giá trị độ không thuộc (ν) vào độ đo khoảng cách Hamming và Euclidean chuẩn hóa như sau:

- Độ đo khoảng cách Hamming trực cảm ($p=1$) và Euclidean trực cảm ($p=2$)

$$d(A, B) = \left(\frac{1}{2} \sum_{i=1}^N \left(|\mu_A(x_i) - \mu_B(x_i)|^p + |\nu_A(x_i) - \nu_B(x_i)|^p \right) \right)^{1/p} \quad (3.4)$$

- Độ đo khoảng cách Hamming ($p=1$) và Euclidean ($p=2$) trực cảm chuẩn hóa

$$d(A, B) = \left(\frac{1}{2N} \sum_{i=1}^N \left(|\mu_A(x_i) - \mu_B(x_i)|^p + |\nu_A(x_i) - \nu_B(x_i)|^p \right) \right)^{1/p} \quad (3.5)$$

Chen [2] and Grzegorzewski [6] đã sửa đổi công thức (3.4 – 3.5) và đưa ra độ đo khoảng cách Hausdroff trực cảm như sau:

- Độ đo khoảng cách Hamming-Hausdorff ($p=1$) và Euclidean-Hausdorff ($p=2$) trực cảm.

$$d(A, B) = \left(\sum_{i=1}^N \max \left\{ |\mu_A(x_i) - \mu_B(x_i)|^p, |\nu_A(x_i) - \nu_B(x_i)|^p \right\} \right)^{1/p} \quad (3.6)$$

- Độ đo khoảng cách Hamming-Hausdorff ($p=1$) và Euclidean-Hausdorff ($p=2$) trực cảm chuẩn hóa.

$$d(A, B) = \left(\frac{1}{N} \sum_{i=1}^N \max \left\{ |\mu_A(x_i) - \mu_B(x_i)|^p, |\nu_A(x_i) - \nu_B(x_i)|^p \right\} \right)^{1/p} \quad (3.7)$$

Một số tác giả đã thử lồng thêm các chỉ số “do dự” vào các công thức (3.4–3.7) và thay đổi hệ số của các giá trị độ thuộc (μ) và độ không thuộc (ν). Tuy nhiên, do hạn chế về độ phức tạp tính toán nên công thức (3.4 – 3.7) vẫn là những độ đo được sử dụng rộng rãi nhất cho IFS. Các công thức tổng quát của các độ đo trên FS là (3.1–3.3)

Mặc dù một số độ đo khoảng cách trên IFS đã có sẵn, nhưng vẫn không thể sử dụng được để tính khoảng cách giữa các tập PFS do có sự tham gia của giá trị độ trung lập (η) trong PFS. Cuong & Kreinovich [3] đã trình bày một số độ đo khoảng cách viễn cảnh trong công thức (2.41 – 2.42), đây là nâng cao của độ đo khoảng cách Hamming và Euclidean trực cảm chuẩn hóa trong công thức (3.5)

Cùng ý tưởng với Xu & Xia [23], trong phần này sẽ tổng quát hóa các độ đo khoảng cách viễn cảnh của Cuong & Kreinovich trong các công thức (2.41 – 2.42), mở đầu bằng định nghĩa 3.1 sau đây

3.1.2. Định nghĩa

Định nghĩa 3.1: Độ đo khoảng cách giữa hai tập mờ viễn cảnh $A, B \in \text{PFS}(X)$, ký hiệu là $d(A, B)$ thỏa mãn các tính chất trong công thức (3.8 – 3.10) như sau:

$$0 \leq d(A, B) \leq 1, \quad (3.8)$$

$$d(A, B) = 0 \Leftrightarrow A = B, \quad (3.9)$$

$$d(A, B) = d(B, A), \quad (3.10)$$

Từ đó, ta sẽ mở rộng độ đo khoảng cách trực cảm trong các công thức (3.4–3.7) trên tập mờ viễn cảnh PFS.

- Độ đo khoảng cách Hamming viễn cảnh ($p=1$) và Euclidean viễn cảnh ($p=2$)

$$d(A, B) = \left(\frac{1}{3} \sum_{i=1}^N \left(|\mu_A(x_i) - \mu_B(x_i)|^p + |\eta_A(x_i) - \eta_B(x_i)|^p + |\nu_A(x_i) - \nu_B(x_i)|^p \right) \right)^{1/p} \quad (3.11)$$

- Độ đo khoảng cách Hamming ($p=1$) và Euclidean ($p=2$) viễn cảnh chuẩn hóa

$$d(A, B) = \left(\frac{1}{3N} \sum_{i=1}^N \left(|\mu_A(x_i) - \mu_B(x_i)|^p + |\eta_A(x_i) - \eta_B(x_i)|^p + |\nu_A(x_i) - \nu_B(x_i)|^p \right) \right)^{1/p} \quad (3.12)$$

- Độ đo khoảng cách Hamming-Hausdorff (p=1) và Euclidean-Hausdorff (p=2) viễn cảnh

$$d(A, B) = \left(\sum_{i=1}^N \max \left\{ |\mu_A(x_i) - \mu_B(x_i)|^p, |\eta_A(x_i) - \eta_B(x_i)|^p, |\nu_A(x_i) - \nu_B(x_i)|^p \right\} \right)^{1/p} \quad (3.13)$$

- Độ đo khoảng cách Hamming-Hausdorff (p=1) và Euclidean-Hausdorff (p=2) viễn cảnh chuẩn hóa

$$d(A, B) = \left(\frac{1}{N} \sum_{i=1}^N \max \left\{ |\mu_A(x_i) - \mu_B(x_i)|^p, |\eta_A(x_i) - \eta_B(x_i)|^p, |\nu_A(x_i) - \nu_B(x_i)|^p \right\} \right)^{1/p} \quad (3.14)$$

Ta có thể thấy rằng, các công thức (3.11 – 3.14) là tổng quát hóa của (3.4 – 3.7), và các độ đo khoảng cách của Cuong & Kreinovich là một trường hợp đặc biệt của công thức (3.11–3.14). Để đơn giản hóa, ta ký hiệu:

$$\Delta\mu_i = |\mu_A(x_i) - \mu_B(x_i)|, (i = \overline{1, N}) \quad (3.15)$$

$$\Delta\eta_i = |\eta_A(x_i) - \eta_B(x_i)|, (i = \overline{1, N}) \quad (3.16)$$

$$\Delta\nu_i = |\nu_A(x_i) - \nu_B(x_i)|, (i = \overline{1, N}) \quad (3.17)$$

Như vậy, ta sẽ có độ đo khoảng cách lai ghép giữa độ đo trong (3.11, 3.13) để có được độ đo khoảng cách Hamming-Hausdorff viễn cảnh lai ghép (p=1) và độ đo khoảng cách Euclidean-Hausdorff viễn cảnh lai ghép (p=2) như sau:

$$d(A, B) = \left(\sum_{i=1}^N \left(\frac{\Delta\mu_i^p + \Delta\eta_i^p + \Delta\nu_i^p}{3} + \max \{ \Delta\mu_i^p, \Delta\eta_i^p, \Delta\nu_i^p \} \right) \right)^{1/p} \quad (3.18)$$

Tương tự, lai ghép giữa độ đo trong (3.12 – 3.14) ta sẽ có độ đo khoảng cách Hamming-Hausdorff viễn cảnh chuẩn hóa lai ghép (p=1) và độ đo khoảng cách Euclidean-Hausdorff viễn cảnh chuẩn hóa lai ghép (p=2):

$$d(A, B) = \left(\frac{1}{N} \sum_{i=1}^N \left(\frac{\Delta\mu_i^p + \Delta\eta_i^p + \Delta\nu_i^p}{3} + \max \{ \Delta\mu_i^p, \Delta\eta_i^p, \Delta\nu_i^p \} \right) \right)^{1/p} \quad (3.19)$$

Đến đây, ta định nghĩa độ đo khoảng cách Hamming viễn cảnh (p=1) và Euclidean viễn cảnh (p=2) tổng quát như sau:

$$d(A, B) = \left(\sum_{i=1}^N |(\mu_A(x_i) + \eta_A(x_i) + \nu_A(x_i)) - (\mu_B(x_i) + \eta_B(x_i) + \nu_B(x_i))|^p \right)^{1/p} \quad (3.20)$$

Độ đo khoảng cách Hamming viễn cảnh chuẩn hóa tổng quát (p=1) và độ đo khoảng cách Euclidean viễn cảnh chuẩn hóa tổng quát (p=2):

$$d(A, B) = \left(\frac{1}{N} \sum_{i=1}^N |(\mu_A(x_i) + \eta_A(x_i) + \nu_A(x_i)) - (\mu_B(x_i) + \eta_B(x_i) + \nu_B(x_i))|^p \right)^{1/p} \quad (3.21)$$

Độ đo khoảng cách Hausdorff viễn cảnh tổng quát là:

$$d(A, B) = \max_i \{(\mu_A(x_i) + \eta_A(x_i) + \upsilon_A(x_i)), (\mu_B(x_i) + \eta_B(x_i) + \upsilon_B(x_i))\} \quad (3.22)$$

Ta thấy rằng, các công thức (3.20 – 3.22) tương ứng là tổng quát hóa của (3.1 – 3.3). Ta ký hiệu như sau:

$$\Phi_i^A = |\mu_A(x_i) + \eta_A(x_i) + \upsilon_A(x_i)|, (i = \overline{1, N}) \quad (3.23)$$

$$\Phi_i^B = |\mu_B(x_i) + \eta_B(x_i) + \upsilon_B(x_i)|, (i = \overline{1, N}) \quad (3.24)$$

Khi đó, ta có độ đo khoảng cách Hausdorff-Hamming viễn cảnh tổng quát lai ghép (p=1) và độ đo khoảng cách Hausdorff-Euclidean viễn cảnh tổng quát lai ghép (p=2) như sau:

$$d(A, B) = \left(\max_i \{\Phi_i^A, \Phi_i^B\} + \sum_{i=1}^N |\Phi_i^A - \Phi_i^B|^p \right)^{1/p} \quad (3.25)$$

Độ đo khoảng cách Hausdorff-Hamming viễn cảnh tổng quát chuẩn hóa lai ghép (p=1) và độ đo khoảng cách Hausdorff-Euclidean viễn cảnh tổng quát chuẩn hóa lai ghép (p=2) là:

$$d(A, B) = \left(\max_i \{\Phi_i^A, \Phi_i^B\} + \frac{1}{N} \sum_{i=1}^N |\Phi_i^A - \Phi_i^B|^p \right)^{1/p} \quad (3.26)$$

Từ công thức (3.18, 3.25), ta rút ra công thức độ đo khoảng cách Hausdorff-Hamming viễn cảnh tổng quát (p=1) và Hausdorff-Euclidean viễn cảnh tổng quát (p=2) như sau:

$$d(A, B) = \frac{\left(\sum_{i=1}^N \left(\frac{\Delta\mu_i^p + \Delta\eta_i^p + \Delta\upsilon_i^p}{3} + \max\{\Delta\mu_i^p, \Delta\eta_i^p, \Delta\upsilon_i^p\} \right) \right)^{1/p}}{\left(\sum_{i=1}^N \left(\frac{\Delta\mu_i^p + \Delta\eta_i^p + \Delta\upsilon_i^p}{3} + \max\{\Delta\mu_i^p, \Delta\eta_i^p, \Delta\upsilon_i^p\} \right) \right)^{1/p} + \left(\max_i \{\Phi_i^A, \Phi_i^B\} + \sum_{i=1}^N |\Phi_i^A - \Phi_i^B|^p \right)^{1/p} + 1} \quad (3.27)$$

Và độ đo khoảng cách Hausdorff-Hamming viễn cảnh chuẩn hóa tổng quát (p=1) và Hausdorff-Euclidean viễn cảnh chuẩn hóa tổng quát (p=2) như sau:

$$d(A, B) = \frac{\left(\frac{1}{N} \sum_{i=1}^N \left(\frac{\Delta\mu_i^p + \Delta\eta_i^p + \Delta\upsilon_i^p}{3} + \max\{\Delta\mu_i^p, \Delta\eta_i^p, \Delta\upsilon_i^p\} \right) \right)^{1/p}}{\left(\frac{1}{N} \sum_{i=1}^N \left(\frac{\Delta\mu_i^p + \Delta\eta_i^p + \Delta\upsilon_i^p}{3} + \max\{\Delta\mu_i^p, \Delta\eta_i^p, \Delta\upsilon_i^p\} \right) \right)^{1/p} + \left(\max_i \{\Phi_i^A, \Phi_i^B\} + \frac{1}{N} \sum_{i=1}^N |\Phi_i^A - \Phi_i^B|^p \right)^{1/p} + 1} \quad (3.28)$$

Đối với dữ liệu liên tục trên khoảng cho trước như $[a, b]$ thì độ đo khoảng cách viễn cảnh tổng quát (3.27, 3.28) được viết như sau:

$$d(A, B) = \frac{\left(\int_a^b \left(\frac{\Delta\mu_i^p + \Delta\eta_i^p + \Delta\upsilon_i^p}{3} + \max\{\Delta\mu_i^p, \Delta\eta_i^p, \Delta\upsilon_i^p\} \right) \right)^{1/p}}{\left(\int_a^b \left(\frac{\Delta\mu_i^p + \Delta\eta_i^p + \Delta\upsilon_i^p}{3} + \max\{\Delta\mu_i^p, \Delta\eta_i^p, \Delta\upsilon_i^p\} \right) \right)^{1/p} + \left(\max_i \{\Phi_i^A, \Phi_i^B\} + \int_a^b |\Phi_i^A - \Phi_i^B|^p \right)^{1/p} + 1} \quad (3.29)$$

$$d(A, B) = \frac{\left(\frac{1}{b-a} \int_a^b \left(\frac{\Delta\mu_i^p + \Delta\eta_i^p + \Delta\nu_i^p}{3} + \max\{\Delta\mu_i^p, \Delta\eta_i^p, \Delta\nu_i^p\} \right) \right)^{1/p}}{\left(\frac{1}{b-a} \int_a^b \left(\frac{\Delta\mu_i^p + \Delta\eta_i^p + \Delta\nu_i^p}{3} + \max\{\Delta\mu_i^p, \Delta\eta_i^p, \Delta\nu_i^p\} \right) \right)^{1/p} + \left(\max_i \{\Phi_i^A, \Phi_i^B\} + \frac{1}{b-a} \int_a^b |\Phi_i^A - \Phi_i^B|^p \right)^{1/p}} + 1 \quad (3.30)$$

3.1.3. Các định lý

Từ các định nghĩa độ đo khoảng cách viễn cảnh tổng quát trong các công thức (3.27 – 3.30), ta có thể đưa ra được các kết quả sau:

Định lý 1: Các độ đo khoảng cách trong các công thức (3.27–3.30) là các đại lượng đo khoảng cách viễn cảnh tổng quát.

Định lý 2: Các độ đo khoảng cách viễn cảnh tổng quát trong các công thức (3.27–3.30) nhỏ hơn các độ đo của Cuong & Kreinovich trong các công thức (2.41 – 2.42)

3.2. Thuật toán phân cụm phân cấp mờ viễn cảnh

Sau khi lý thuyết tập mờ ra đời, nhiều tác giả đã tiến hành nghiên cứu phân cụm trên tập mờ và phân cụm trên các mở rộng của tập mờ, kết quả thu được là rất nhiều thuật toán phân cụm đã được phát triển như thuật toán phân cụm Fuzzy C-Means (Fan et al. 2004), thuật toán phân cụm cây tối đa (Christopher và Burges 1998), v.v.

Xu (2009) [24] đã giới thiệu thuật toán phân cụm phân cấp mờ trực cảm IHC, thực hiện phân cụm trên tập dữ liệu mờ trực cảm IFS. Thuật toán IHC dựa trên phương pháp phân cụm phân cấp truyền thống (Anderberg 1972), các phép toán tập hợp trên tập mờ trực cảm và các độ đo khoảng cách cơ bản giữa các tập mờ trực cảm IFS (như khoảng cách Hamming, Hamming chuẩn hóa, khoảng cách Euclidean, Euclidean chuẩn hóa, v.v). Tuy nhiên, thuật toán của Xu vẫn không thể sử dụng để thực hiện phân cụm trên tập mờ viễn cảnh PFS do trong PFS có thêm giá trị độ trung lập (η).

Cùng ý tưởng với Xu, trong phần này, ta sẽ sử dụng các độ đo khoảng cách viễn cảnh tổng quát giữa các tập mờ viễn cảnh PFS trong các công thức (3.27 – 3.28) để đề xuất thuật toán phân cụm phân cấp mờ viễn cảnh mới HPC. Thuật toán HPC dựa trên phương pháp phân cụm phân cấp truyền thống, các phép toán trên tập mờ viễn cảnh và các độ đo khoảng cách viễn cảnh tổng quát để thực hiện phân cụm trên tập dữ liệu mờ viễn cảnh PFS. Bắt đầu bằng định nghĩa 3.2 sau:

3.2.1. Định nghĩa giá trị trung bình

Định nghĩa 3.2: Giá trị trung bình của các tập mờ viễn cảnh $A_j \in \text{PFS}(X)$ ($j = \overline{1, M}$), ký hiệu là $\text{AVG}(A_j)$, được xác định như sau:

$$\text{AVG}(A_j) = \left\{ \left\langle x, \frac{1}{M} \sum_{j=1}^M \mu_j(x), \frac{1}{M} \sum_{j=1}^M \eta_j(x), \frac{1}{M} \sum_{j=1}^M \nu_j(x) \right\rangle \mid x \in X \right\} \quad (3.31)$$

với $\mu_j(x), \eta_j(x), \nu_j(x)$ tương ứng là các độ thuộc, độ trung lập và độ không thuộc của x trên tập A_j .

3.2.2. Thuật toán HPC

Input: Cho tập thuộc tính $X=\{x_i\}$, ($i=\overline{1,N}$) và tập các đối tượng là các tập mờ viễn cảnh PFS A_j với $A_j = \{(x_i, \mu_{A_j}(x_i), \eta_{A_j}(x_i), \nu_{A_j}(x_i)) \mid x_i \in X\}$, ($j=\overline{1,M}$).

Output: Các cụm mờ viễn cảnh.

Xử lý: Thuật toán HPC được mô tả qua các bước sau:

Bước 1: Đưa các tập mờ viễn cảnh $A_j \in \text{PFS}(X)$ ($j=\overline{1,M}$) là các cụm riêng biệt.

Bước 2: Tính ma trận khoảng cách viễn cảnh PDM dựa trên các độ đo khoảng cách viễn cảnh tổng quát theo công thức (3.27 – 3.28).

Bước 3: Dựa trên ma trận PDM, thực hiện kết hợp hai cụm liền kề nhau (mỗi lần chỉ kết hợp 2 cụm với nhau và không được tách ra sau khi đã kết hợp cụm). Sau đó, tính toán ra tâm mới theo công thức (3.31).

Bước 4: Với A_j được thay thế bởi trọng tâm mới, thực hiện lặp lại từ bước 2 tới bước 4 cho đến khi đạt được số cụm như mong muốn.

3.3. Đánh giá

3.3.1. Thiết kế thực nghiệm

(1) Công cụ thực nghiệm

- Với các độ đo khoảng cách viễn cảnh tổng quát trong công thức (3.27 – 3.28) sẽ thu được thuật toán HPC với bốn phiên bản HPC1, HPC2, HPC3, HPC4 như sau:
 - Thuật toán HPC1 sử dụng độ đo khoảng cách Hausdorff-Hamming viễn cảnh tổng quát ($p=1$) trong công thức (3.27).
 - Thuật toán HPC2 sử dụng độ đo khoảng cách Hausdorff-Euclidean viễn cảnh tổng quát ($p=2$) trong công thức (3.27).
 - Thuật toán HPC3 sử dụng độ đo khoảng cách Hausdorff-Hamming viễn cảnh chuẩn hóa tổng quát ($p=1$) trong công thức (3.28).
 - Thuật toán HPC4 sử dụng độ đo khoảng cách Hausdorff- Euclidean viễn cảnh chuẩn hóa tổng quát ($p=2$) trong công thức (3.28).
- Với các độ đo khoảng cách viễn cảnh của Cuong & Kreinovich sẽ thu được thuật toán CK với hai phiên bản CK1, CK2 như sau:
 - Thuật toán CK1 sử dụng độ đo khoảng cách Hamming chuẩn hóa trong công thức (2.41) của Cuong & Kreinovich [3].
 - Thuật toán CK2 sử dụng độ đo khoảng cách Euclidean chuẩn hóa trong công thức (2.42) của Cuong & Kreinovich [3].
- Với độ đo khoảng cách giữa các tập mờ trực cảm IFS [24] (bằng cách kết hợp hai giá trị độ trung lập và độ không thuộc trên tập PFS) sẽ thu được thuật toán IHC như sau:
 - Thuật toán IHC trong tài liệu [24] được sử dụng để đánh giá chất lượng phân cụm.

(2) Dữ liệu thực nghiệm

Thực nghiệm thuật toán HPC sẽ sử dụng ba bộ dữ liệu bao gồm: bộ dữ liệu xe hơi Quảng Châu [30] được biểu diễn như bảng 3.1, bộ dữ liệu các vật liệu xây dựng [21, 24] được biểu diễn ở bảng 3.2 và một bộ dữ liệu thực tế liên quan đến chẩn đoán bệnh tim được lấy từ UCI Machine Learning Repository [18].

Trong bảng 3.1, có 5 loại xe hơi mới ra đời được xếp loại vào thị trường xe hơi Quảng Châu ở Quảng Đông, Trung Quốc, và 6 thuộc tính: tiết kiệm nhiên liệu (G1), đẳng cấp (G2), giá cả (G3), tiện nghi (G4), thiết kế (G5), an toàn (G6) được đem ra xem xét khi tiến hành xếp loại. Các đặc điểm của 5 loại xe hơi mới theo 6 thuộc tính được biểu diễn như sau:

Bảng 3.1. Bộ dữ liệu xe hơi Quảng Châu

	Fuel (G1)	Aerod (G2)	Price (G3)
Car1	(0.3,0.1,0.4)	(0.6,0.05,0.05)	(0.4,0.1,0.2)
Car2	(0.6,0.2,0.1)	(0.5,0.1,0.1)	(0.6,0,0.1)
Car3	(0.4,0.3,0.2)	(0.1,0.5,0.3)	(0.1,0.2,0.4)
Car4	(0.2,0.2,0.2)	(0.4,0.05,0.05)	(0.9,0,0)
Car5	(0,0.4,0.1)	(0.4,0.1,0.5)	(0.1,0.6,0.2)
	Comfort (G4)	Design (G5)	Safety (G6)
Car1	(0.8,0,0.1)	(0.1,0.2,0.4)	(0.5,0.3,0.1)
Car2	(0.1,0.1,0)	(0.3,0.3,0.3)	(0.4,0.1,0.2)
Car3	(0.2,0.1,0.5)	(0.3,0.4,0.2)	(0.6,0.3,0.1)
Car4	(0.8,0.1,0)	(0.2,0.2,0.3)	(0.7,0.03,0.07)
Car5	(0.1,0.1,0)	(0.6,0.1,0.1)	(0.2,0.1,0.2)

Loại xe hơi Car1 theo thuộc tính tiết kiệm nhiên liệu (G1) là (0.3, 0.1, 0.4), có nghĩa là $\mu_{\text{Car1}}(G1) = 0.3$, $\eta_{\text{Car1}}(G1) = 0.1$ và $\nu_{\text{Car1}}(G1) = 0.4$ thỏa mãn điều kiện:

$$(\mu_{\text{Car1}}(G1) + \eta_{\text{Car1}}(G1) + \nu_{\text{Car1}}(G1)) \leq 1.$$

Tương tự với các loại xe hơi khác theo các thuộc tính còn lại.

Bảng 3.2 gồm có 5 loại vật liệu xây dựng: keo (Sealant), véc ni (Varnish), sơn tường (Wall paint), thảm (Carpet), ván sàn nhựa PVC (Chloride flooring) trong không gian thuộc tính $X = \{X_1, X_2, \dots, X_8\}$.

Đối với vật liệu sơn tường (Wall paint) theo thuộc tính X_1 là (0.45, 0.1, 0.25), có nghĩa là $\mu_{\text{Wallpaint}}(X_1) = 0.45$, $\eta_{\text{Wallpaint}}(X_1) = 0.1$ và $\nu_{\text{Wallpaint}}(X_1) = 0.25$ thỏa mãn điều kiện:

$$(\mu_{\text{Wallpaint}}(X_1) + \eta_{\text{Wallpaint}}(X_1) + \nu_{\text{Wallpaint}}(X_1)) \leq 1$$

Tương tự với các vật liệu khác theo các thuộc tính còn lại.

Bảng 3.2. Bộ dữ liệu các vật liệu xây dựng

	X₁	X₂	X₃	X₄
Sealant	(0.9,0,0)	(0.1,0.3,0.5)	(0.5,0.1,0.2)	(0.2,0,0)
Floor varnish	(0.5,0.2,0.2)	(0.6,0.1,0.05)	(1,0,0)	(0.15,0.3,0.35)
Wall paint	(0.45,0.1,0.25)	(0.6,0.1,0.2)	(0.9,0,0)	(0.1,0.5,0.3)
Carpet	(1,0,0)	(1,0,0)	(0.85,0.05,0.05)	(0.75,0.15,0)
Chloride flooring	(0.9,0,0)	(0.9,0.1,0)	(0.8,0,0.1)	(0.7,0.1,0.1)
	X₅	X₆	X₇	X₈
Sealant	(0.4,0.15,0.2)	(0.1,0.4,0.5)	(0.3,0.3,0.2)	(0.5,0.1,0)
Floor varnish	(0,0.3,0.5)	(0.7,0.05,0.1)	(0.5,0.1,0.2)	(0.65,0.1,0.1)
Wall paint	(0.2,0.3,0.4)	(0.6,0.1,0.1)	(0.15,0.4,0.4)	(0.2,0.05,0.6)
Carpet	(0.2,0.2,0.6)	(0.15,0.25,0.6)	(0.1,0.3,0.4)	(0.3,0.3,0.4)
Chloride flooring	(0.5,0.05,0.1)	(0.3,0.35,0.3)	(0.15,0.25,0.5)	(0.4,0.2,0.1)

Đối với dữ liệu bệnh Tim lấy từ [18], sử dụng bộ dữ liệu được thu thập từ Cleveland Clinic Foundation gồm 270 bản ghi và 3 thuộc tính quan trọng là: tuổi (3#Age), huyết áp (mm Hg trên bệnh nhân nhập viện) (10# Trestbps) và nhịp tim tối đa đạt được (#32 Thalach). Một phần dữ liệu thực nghiệm này được biểu diễn như bảng 3.3 sau:

Bảng 3.3. Một phần bộ dữ liệu bệnh tim từ UCI

ID	Age	Trestbps	Thalach
X1	(0,0.292,0.438)	(0,0.679,0.321)	(0,0.5,0.5)
X2	(0,0.417,0.583)	(0,0.396,0.604)	(0,0.5,0.5)
X3	(0,0.833,0.167)	(0,0.566,0.434)	(0,0.5,0.5)
X4	(0,0.542,0.458)	(0,0.642,0.358)	(0,0.5,0.5)
X5	(0,0.125,0.188)	(0,0.491,0.509)	(0,0.5,0.5)
X6	(0,0.5,0.5)	(0,0.491,0.509)	(0,0.5,0.5)
X7	(0.011,0.875,0.114)	(0,0.679,0.321)	(0,0.5,0.5)
X8	(0,0.75,0.25)	(0,0.302,0.453)	(0,0.5,0.5)
X9	(0,0.708,0.292)	(0,0.868,0.132)	(0,0.5,0.5)
X10	(0,0.583,0.417)	(0.011,0.943,0.046)	(0,0.5,0.5)
X11	(0,0.75,0.25)	(0,0.774,0.226)	(0,0.5,0.5)
X12	(1,0,0)	(0,0.906,0.094)	(0,0.5,0.5)
X13	(0,0.625,0.375)	(0,0.868,0.132)	(0,0.5,0.5)
X14	(0,0.667,0.333)	(0,0.755,0.245)	(0,0.5,0.5)
X15	(0,0.833,0.167)	(0,0.642,0.358)	(0,0.5,0.5)

(3) **Đánh giá chất lượng phân cụm**

Để thực hiện đánh giá chất lượng phân cụm của CK1, CK2, HPC1, HPC2, HPC3, HPC4, ta sử dụng kết quả phân cụm của IHC làm chuẩn cơ sở, xác định một số chỉ số đánh giá như là NMI (Normalized Mutual Information), F-Measure và Purity. Giá trị các chỉ số đánh giá càng cao có nghĩa là chất lượng phân cụm càng tốt.

a. **NMI (Normalized Mutual Information)**

Chỉ số NMI được xác định như sau:

$$NMI = \frac{\sum_{j=1}^k \sum_{i=1}^r n_{ij} \log \frac{n \cdot n_{ij}}{n_i \cdot n_j}}{\sqrt{(\sum_{i=1}^r n_i \log \frac{n_i}{n})(\sum_{j=1}^k n_j \log \frac{n_j}{n})}} \quad (3.32)$$

Trong đó:

- $T=\{T_1, \dots, T_k\}$: cụm chuẩn gồm k cụm
- $C=\{C_1, \dots, C_r\}$: cụm gồm r cụm
- n: tổng số lượng điểm
- $n_{ij} = |C_i \cap T_j|$: số lượng điểm chung của cụm C_i và T_j
- $n_j = \sum_{i=1}^r n_{ij}$: số lượng điểm trong cụm T_j
- $n_i = \sum_{j=1}^k n_{ij}$: số lượng điểm trong cụm C_i

b. **F-Measure**

Chỉ số Precision, Recall và F-Measure của cụm C_i được xác định trong các công thức (3.33-3.35) như sau:

$$Precision_i = \frac{1}{n_i} \max_{j=1}^k \{n_{ij}\} \quad (3.33)$$

$$Recall_i = \frac{1}{n_j} \max_{j=1}^k \{n_{ij}\} \quad (3.34)$$

$$F_i = \frac{2 * Precision_i * Recall_i}{Precision_i + Recall_i} \quad (3.35)$$

Chỉ số F-measure tổng quát được xác định trong công thức (3.36) như sau:

$$F = \frac{1}{r} \sum_{i=1}^r F_i \quad (3.36)$$

Trong đó:

- $T=\{T_1, \dots, T_k\}$: cụm chuẩn gồm k cụm

- $C=\{C_1, \dots, C_r\}$: cụm gồm r cụm
- n : tổng số lượng điểm
- $n_{ij} = |C_i \cap T_j|$: số lượng điểm chung của cụm C_i và T_j
- $n_j = \sum_{i=1}^r n_{ij}$: số lượng điểm trong cụm T_j
- $n_i = \sum_{j=1}^k n_{ij}$: số lượng điểm trong cụm C_i

c. Purity

Chỉ số Purity được xác định như sau:

$$\text{Purity} = \frac{1}{n} \sum_{i=1}^r \max_{j=1}^k \{n_{ij}\} \quad (3.37)$$

Trong đó:

- $T=\{T_1, \dots, T_k\}$: cụm chuẩn gồm k cụm
- $C=\{C_1, \dots, C_r\}$: cụm gồm r cụm
- n : tổng số lượng điểm
- $n_{ij} = |C_i \cap T_j|$: số lượng điểm chung của cụm C_i và T_j

(4) Mục đích thực nghiệm

Mục đích của thực nghiệm là minh họa hoạt động của các thuật toán đã đề xuất. Đồng thời, qua thực nghiệm để đánh giá xem liệu chất lượng phân cụm của các thuật toán (HPC1, HPC2, HPC3, HPC4) sử dụng các độ đo khoảng cách viễn cảnh tổng quát có tốt hơn so với các thuật toán sử dụng các độ đo khoảng cách viễn cảnh của Cuong & Kreinovich (CK1, CK2).

3.3.2. Minh họa thuật toán

Trong phần này sẽ minh họa hoạt động của thuật toán HPC2 sử dụng độ đo khoảng cách Hausdroff-Euclidean trên bộ dữ liệu xe hơi Quảng Châu ở bảng 3.1. Hoạt động của các thuật toán còn lại trên các bộ dữ liệu thực nghiệm được thực hiện tương tự.

Ban đầu, mỗi xe hơi trong bộ dữ liệu là một cụm riêng biệt, được biểu diễn như sau:

$$\{\text{Car1}\}, \{\text{Car2}\}, \{\text{Car3}\}, \{\text{Car4}\}, \{\text{Car5}\}$$

Ma trận khoảng cách viễn cảnh PDM được tính toán bằng độ đo khoảng cách Hausdroff-Euclidean viễn cảnh tổng quát như sau:

$$\text{PDM}^1 = \begin{pmatrix} 0 & 0.2908 & 0.3275 & 0.2560 & 0.3406 \\ 0.2908 & 0 & 0.3036 & 0.2912 & 0.3286 \\ 0.3275 & 0.3036 & 0 & 0.3581 & 0.3062 \\ 0.2560 & 0.2912 & 0.3581 & 0 & 0.3767 \\ 0.3406 & 0.3286 & 0.3062 & 0.3767 & 0 \end{pmatrix}$$

Do $d(\text{Car1}, \text{Car4}) = 0.2560$ là giá trị nhỏ nhất trong tất cả các giá trị khoảng cách trong ma trận PDM^1 , nên Car1 và Car4 được nhóm lại thành một cụm. Loại bỏ tất cả các độ đo khoảng cách có liên quan tới Car1 và Car4. Như vậy, chỉ còn những độ đo khoảng cách có liên quan tới Car2, Car3, Car5. Khi đó, ta thấy $d(\text{Car2}, \text{Car3}) = 0.3036$ là giá trị nhỏ nhất so với các giá trị còn lại nên Car2 và Car3 được kết hợp với nhau thành một cụm. Như vậy, kết quả phân cụm giai đoạn 1 như sau:

$$\{\text{Car1}, \text{Car4}\}, \{\text{Car2}, \text{Car3}\}, \{\text{Car5}\}$$

Trong kết quả phân cụm giai đoạn 1 thì Car5 vẫn là một cụm riêng biệt và tâm Car5 (ký hiệu là C_5) không thay đổi. Theo định nghĩa giá trị trung bình trong công thức (3.31), tâm của $\{\text{Car1}, \text{Car4}\}$ (ký hiệu là C_{14}), tâm của $\{\text{Car2}, \text{Car3}\}$ (ký hiệu là C_{23}) được tính toán theo bảng 3.4 sau:

Bảng 3.4. Bảng dữ liệu tâm cụm trong giai đoạn 2 thuật toán HPC2

	Fuel (G1)	Aerod (G2)	Price (G3)
C₁₄	(0.25,0.15,0.3)	(0.5, 0.05, 0.05)	(0.65, 0.05, 0.1)
C₂₃	(0.5, 0.25, 0.15)	(0.3,0.3,0.2)	(0.35,0.1,0.25)
C₅	(0,0.4,0.1)	(0.4,0.1,0.5)	(0.1,0.6,0.2)
	Comfort (G4)	Design (G5)	Safety (G6)
C₁₄	(0.8,0.05,0.05)	(0.15,0.2,0.35)	(0.6,0.165,0.085)
C₂₃	(0.15,0.1,0.25)	(0.3,0.35,0.25)	(0.5,0.2,0.15)
C₅	(0.1,0.1,0)	(0.6,0.1,0.1)	(0.2,0.1,0.2)

Ma trận khoảng cách viễn cảnh PDM của giai đoạn 2 được tính toán như sau:

$$\text{PDM}^2 = \begin{pmatrix} 0 & 0.2887 & 0.3485 \\ 0.2887 & 0 & 0.2958 \\ 0.3485 & 0.2958 & 0 \end{pmatrix}$$

Do $d(\{\text{Car1}, \text{Car4}\}, \{\text{Car2}, \text{Car3}\}) = 0.2887$ là giá trị nhỏ nhất trong tất cả các giá trị khoảng cách đã tính toán trong PDM^2 nên kết hợp vào thành một cụm. Như vậy, kết quả phân cụm giai đoạn 2 là như sau:

$$\{\text{Car1}, \text{Car4}, \text{Car2}, \text{Car3}\}, \{\text{Car5}\}$$

Trong kết quả phân cụm giai đoạn 2 thì Car5 vẫn là một cụm riêng biệt và tâm Car5 (ký hiệu là C_5) không thay đổi. Theo công thức (3.31), tâm của $\{\text{Car1}, \text{Car4}, \text{Car2}, \text{Car3}\}$ (ký hiệu là C_{1423}) được tính toán theo bảng 3.5 sau:

Bảng 3.5. Bảng dữ liệu tâm cụm trong giai đoạn 3 thuật toán HPC2

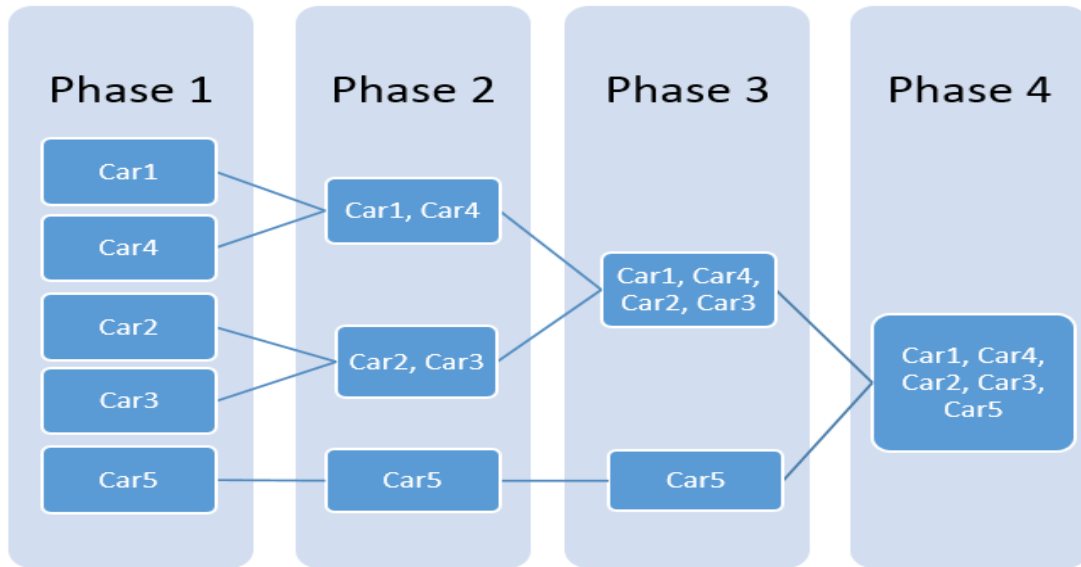
	Fuel (G1)	Aerod (G2)	Price (G3)
C₁₄₂₃	(0.375,0.2,0.225)	(0.4,0.175,0.125)	(0.5,0.075,0.175)
C₅	(0,0.4,0.1)	(0.4,0.1,0.5)	(0.1,0.6,0.2)

	Comfort (G4)	Design (G5)	Safety (G6)
C₁₄₂₃	(0.475,0.075,0.15)	(0.225,0.275,0.3)	(0.55,0.1825,0.1175)
C₅	(0.1,0.1,0)	(0.6,0.1,0.1)	(0.2,0.1,0.2)

PDM của giai đoạn 3 như sau: $PDM^3 = 0.3110$

Cuối cùng, trong giai đoạn 4 tất cả các Car được nhóm vào thành một cụm duy nhất.

Cây phân cấp phân loại bộ dữ liệu xe hơi Quảng Châu sử dụng thuật toán HPC2 được biểu diễn như hình 3.1 sau:



Hình 3.1: Cây phân cấp của thuật toán HPC2 trên dữ liệu xe hơi Quảng Châu

Nếu lấy giá trị trung bình của các độ thuộc, độ trung lập và độ không thuộc thì tâm dữ liệu ở từng giai đoạn của thuật toán HPC2 được thể hiện tương ứng như bảng 3.6 dưới đây

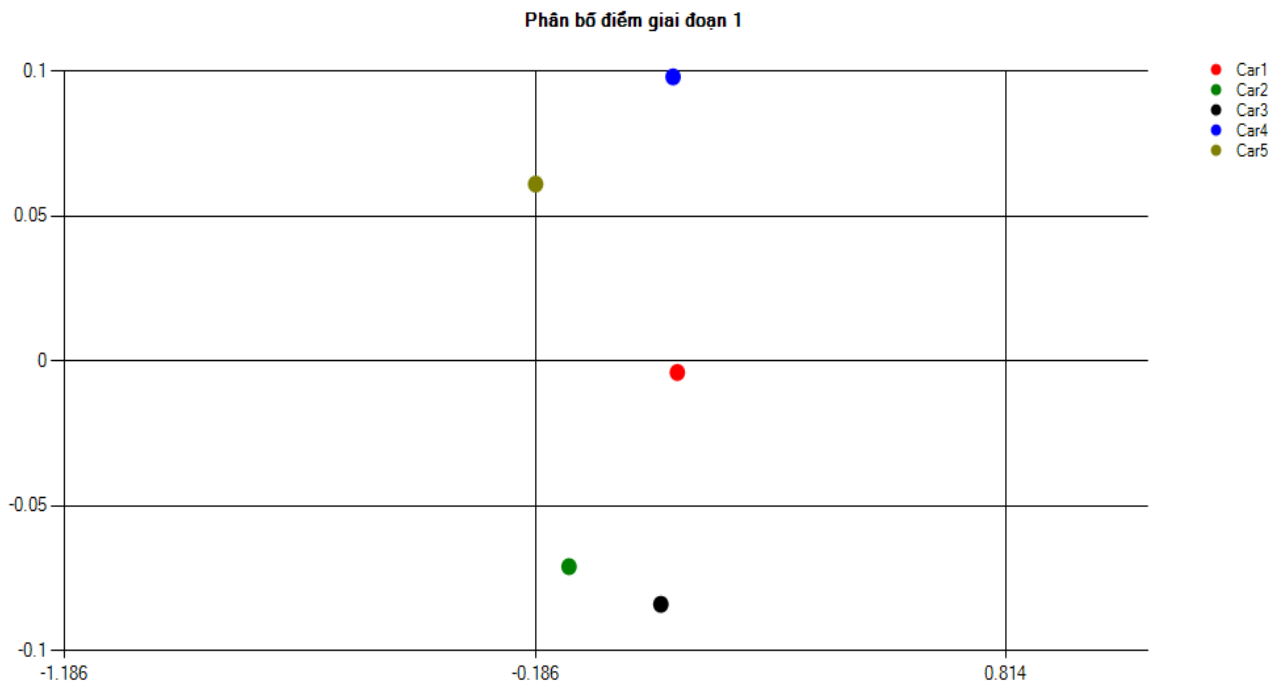
Bảng 3.6. Bảng dữ liệu giá trị trung bình ở các giai đoạn của thuật toán HPC2

ID	Fuel(G₁)	Aerod(G₂)	Price(G₃)	Comfort(G₄)	Design(G₅)	Safety(G₆)
Giai đoạn 1						
Car1	0.267	0.233	0.233	0.3	0.233	0.3
Car4	0.3	0.233	0.233	0.067	0.3	0.233
Car2	0.3	0.3	0.233	0.267	0.3	0.333
Car3	0.2	0.167	0.3	0.3	0.233	0.267
Car5	0.167	0.333	0.3	0.067	0.267	0.167
Giai đoạn 2						
C14	0.233	0.2	0.267	0.3	0.233	0.283
C23	0.3	0.267	0.233	0.167	0.3	0.283
C5	0.167	0.333	0.3	0.067	0.267	0.167
Giai đoạn 3						
C1423	0.267	0.233	0.25	0.233	0.267	0.283
C5	0.167	0.333	0.3	0.067	0.267	0.167

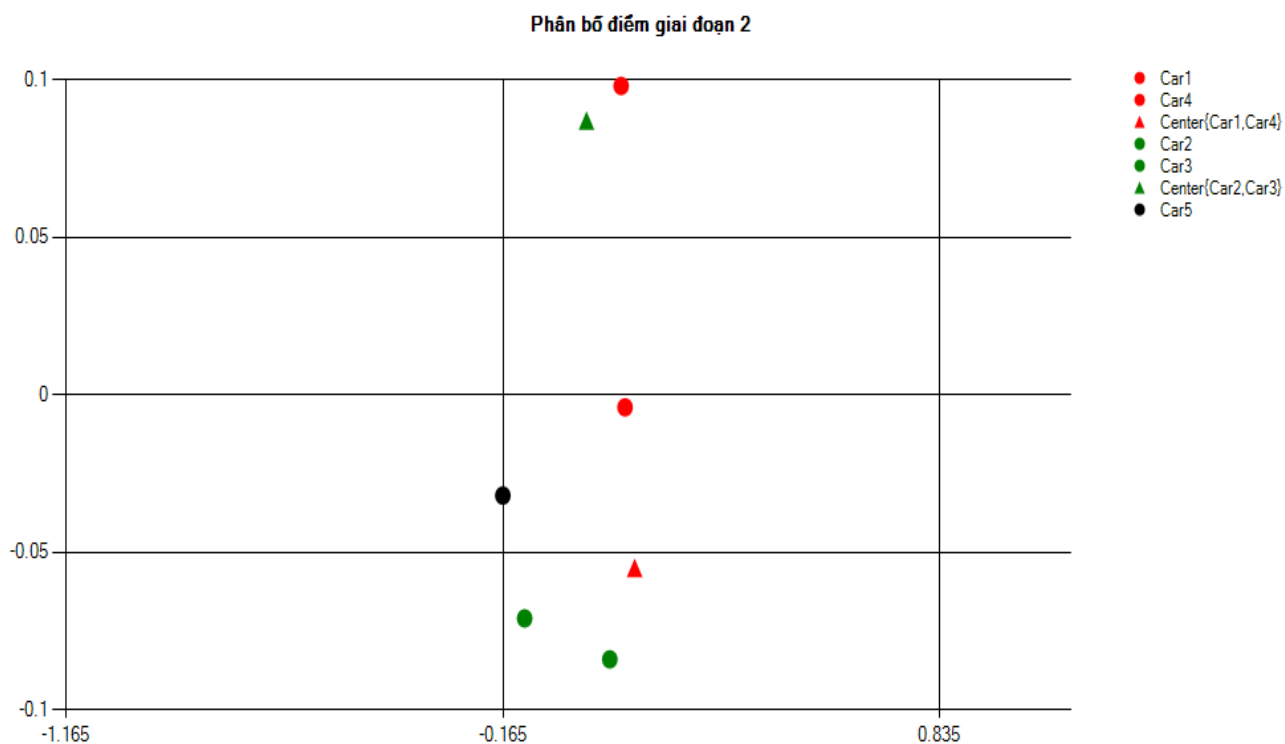
Nếu sử dụng phương pháp phân tích thành phần chính PCA [14] (Principal Components Analysis), một thuật toán thống kê sử dụng phép biến đổi trực giao để biến đổi một tập hợp dữ liệu từ một không gian nhiều chiều sang một không gian mới ít chiều hơn (2 hoặc 3 chiều) nhằm tối ưu hóa việc thể hiện sự biến thiên của dữ liệu, thì bảng dữ liệu 3.6 sau khi áp dụng PCA sẽ thu được kết quả như bảng 3.7 dưới đây. Bảng 3.7. Bảng dữ liệu giá trị trung bình ở các giai đoạn của thuật toán HPC2 sau khi áp dụng PCA

ID	X	Y
Giai đoạn 1		
Car1	0.115	-0.004
Car4	-0.115	-0.072
Car2	0.08	-0.084
Car3	0.106	0.098
Car5	-0.186	0.061
Giai đoạn 2		
C14	0.137	-0.055
C23	0.027	0.087
C5	-0.165	-0.032
Giai đoạn 3		
C1423	0.126	0
C5	-0.126	0

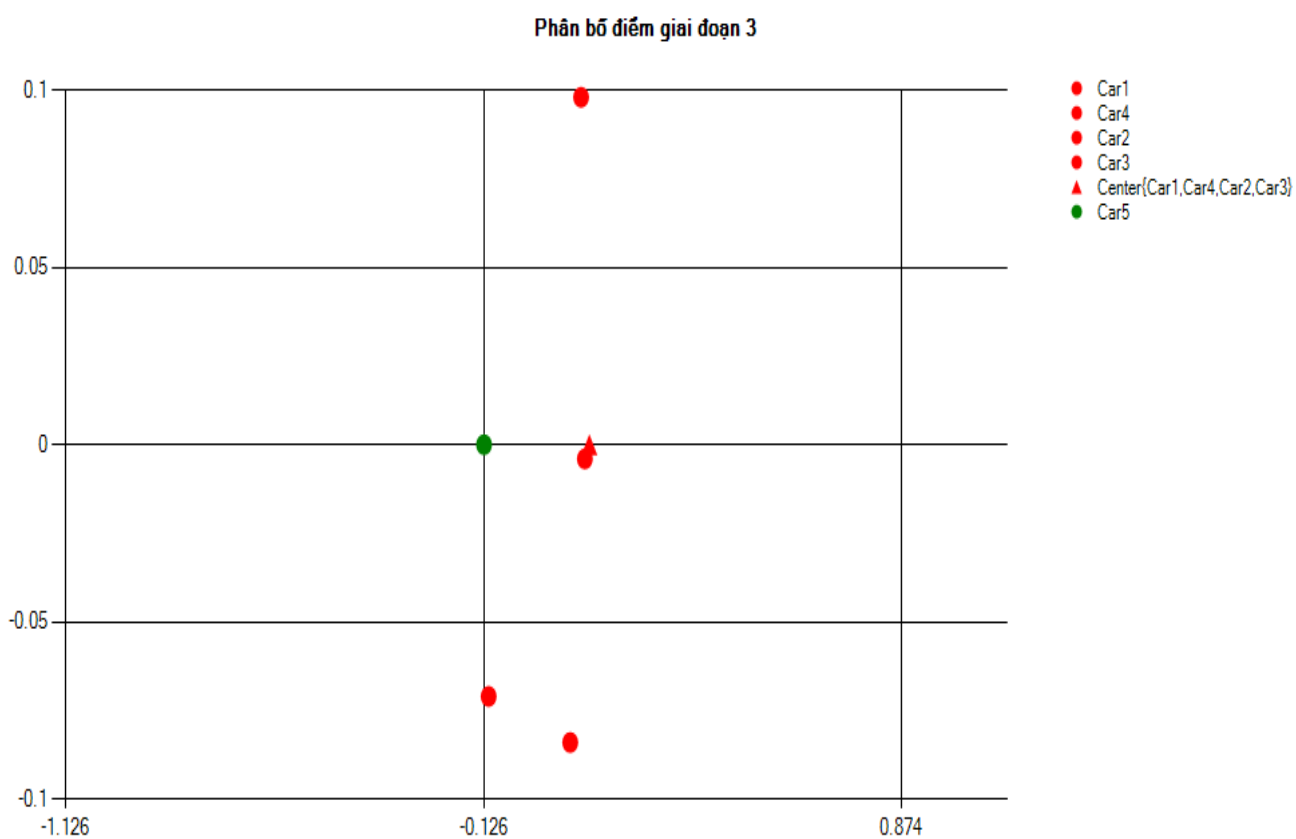
Với dữ liệu 2 chiều ở bảng 3.7 thì phân bố điểm dữ liệu và tâm ở tất cả các giai đoạn của thuật toán HPC2 được minh họa từ hình 3.2 đến hình 3.5 dưới đây



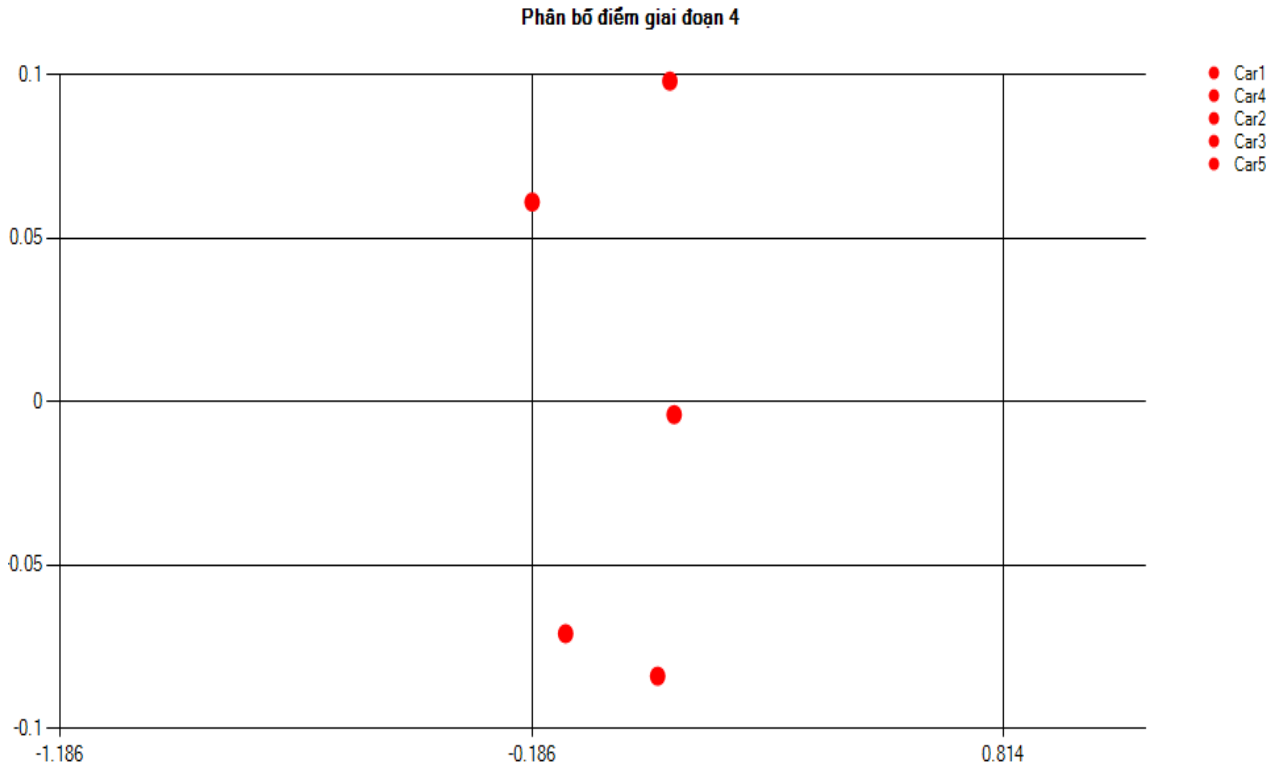
Hình 3.2: Phân bố 2 chiều điểm dữ liệu và tâm trong giai đoạn 1 thuật toán HPC2



Hình 3.3: Phân bố 2 chiều điểm dữ liệu và tâm trong giai đoạn 2 thuật toán HPC2



Hình 3.4: Phân bố 2 chiều điểm dữ liệu và tâm trong giai đoạn 3 thuật toán HPC2



Hình 3.5: Phân bố 2 chiều điểm dữ liệu và tâm trong giai đoạn 4 thuật toán HPC2

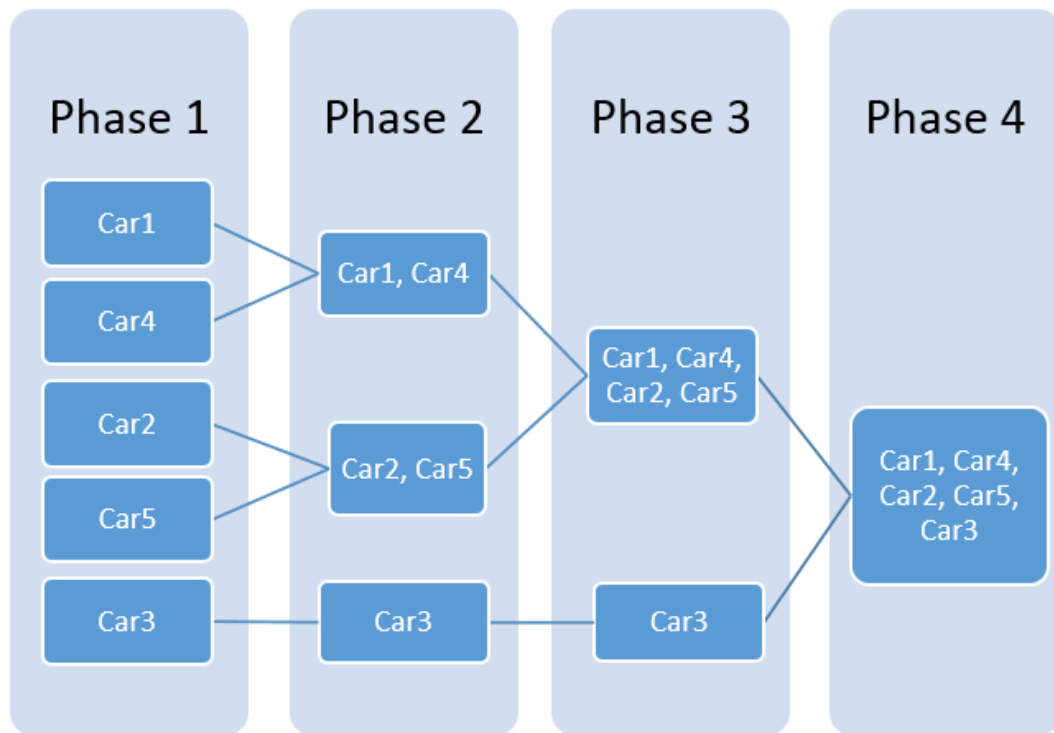
3.3.3. So sánh chất lượng phân cụm

Trong phần này, sẽ đánh giá chất lượng phân cụm của các thuật toán đề xuất (HPC1, HPC2, HPC3, HPC4) so với các thuật toán của Cuong & Kreinovich (CK1, CK2) trên cả ba bộ dữ liệu ở bảng 3.1 – 3.3.

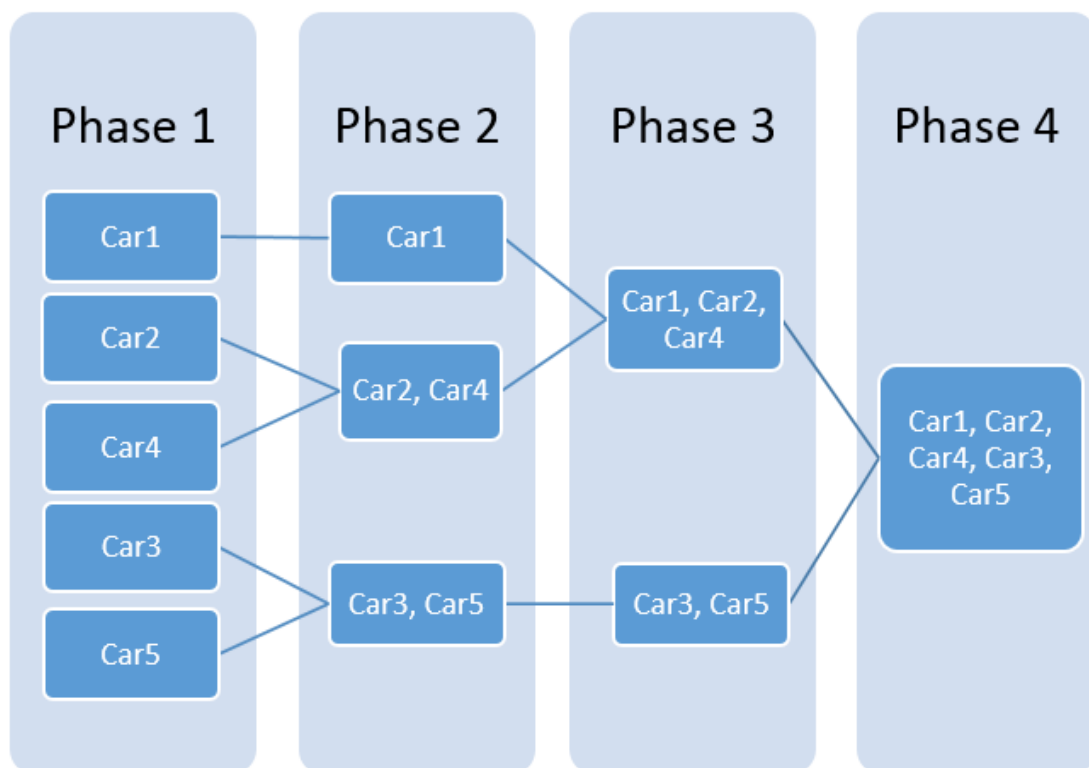
Trong phần 3.3.2 đã minh họa hoạt động của thuật toán HPC2 sử dụng độ đo khoảng cách Hausdroff-Euclidean trên bộ dữ liệu xe hơi Quảng Châu ở bảng 3.1. Bằng cách thực hiện tính toán như tương tự, ta thu được các cây phân cấp của các thuật toán còn lại trên cả ba bộ dữ liệu.

Cây phân cấp của các thuật toán CK2, HPC2, HPC3, HPC4 thực hiện trên bộ dữ liệu ở bảng 3.1 đều giống hệt nhau như ở hình 3.1. Nếu sử dụng IHC, thuật toán phân cụm phân cấp mờ trực cảm tốt nhất trên bộ dữ liệu xe hơi Quảng Châu bằng cách kết hợp hai giá trị độ trung lập và độ không thuộc, thì kết quả cây phân cấp của IHC cũng giống hệt với CK2, HPC2, HPC3 và HPC4 như ở hình 3.1.

Cây phân cấp của thuật toán CK1, HPC1 khác với IHC và tương ứng như trong hình 3.6 và hình 3.7 dưới đây. Trong số hai cây phân cấp trong hình 3.6 và 3.7, cây phân cấp của thuật toán HPC1 (hình 3.7) cung cấp thêm thông tin chi tiết hơn CK1 (hình 3.6) và không có giá trị ngoại lệ (tức là một phần tử thuộc một cụm) khi số lượng các giai đoạn tăng lên.



Hình 3.6: Cây phân cấp của thuật toán CK1 trên bộ dữ liệu xe hơi Quảng Châu



Hình 3.7: Cây phân cấp của thuật toán HPC1 trên bộ dữ liệu xe hơi Quảng Châu

Ta sử dụng IHC làm chuẩn cơ sở, thực hiện tính toán các chỉ số đánh giá gồm NMI, F-Measure và Purity của các thuật toán CK1, CK2, HPC1, HPC2, HPC3, HPC4 (so với IHC) ở từng giai đoạn. Với kết quả phân cụm như ở bảng 3.8 thực hiện trên bộ dữ liệu xe hơi ở bảng 3.1, ta thu được các bảng kết quả chỉ số đánh giá từ bảng 3.9 – 3.11 như dưới đây.

Bảng 3.8. Bảng kết quả phân cụm của IHC, CK1, CK2, HPC1, HPC2, HPC3, HPC4 trên bộ dữ liệu xe hơi Quảng Châu

Giai đoạn	IHC	CK1	HPC1	CK2,HPC2,HPC3,HPC4
2	Car1,Car4	Car1,Car4	Car2,Car4	Car1,Car4
	Car2,Car3	Car2,Car5	Car3,Car5	Car2,Car3
	Car5	Car3	Car1	Car5
3	Car1,Car4,Car2,Car3	Car1,Car4,Car2,Car5	Car2,Car4,Car3,Car5	Car1,Car4, Car2,Car3
	Car5	Car3	Car1	Car5

Bảng 3.9. Bảng chỉ số NMI của CK1, CK2, HPC1, HPC2, HPC3, HPC4 trên bộ dữ liệu xe hơi Quảng Châu

Giai đoạn	CK1	HPC1	CK2, HPC2, HPC3, HPC4
2	0.737	0.474	1
3	0.101	0.385	1

Bảng 3.10. Bảng chỉ số F-Measure của CK1, CK2, HPC1, HPC2, HPC3, HPC4 trên bộ dữ liệu xe hơi Quảng Châu

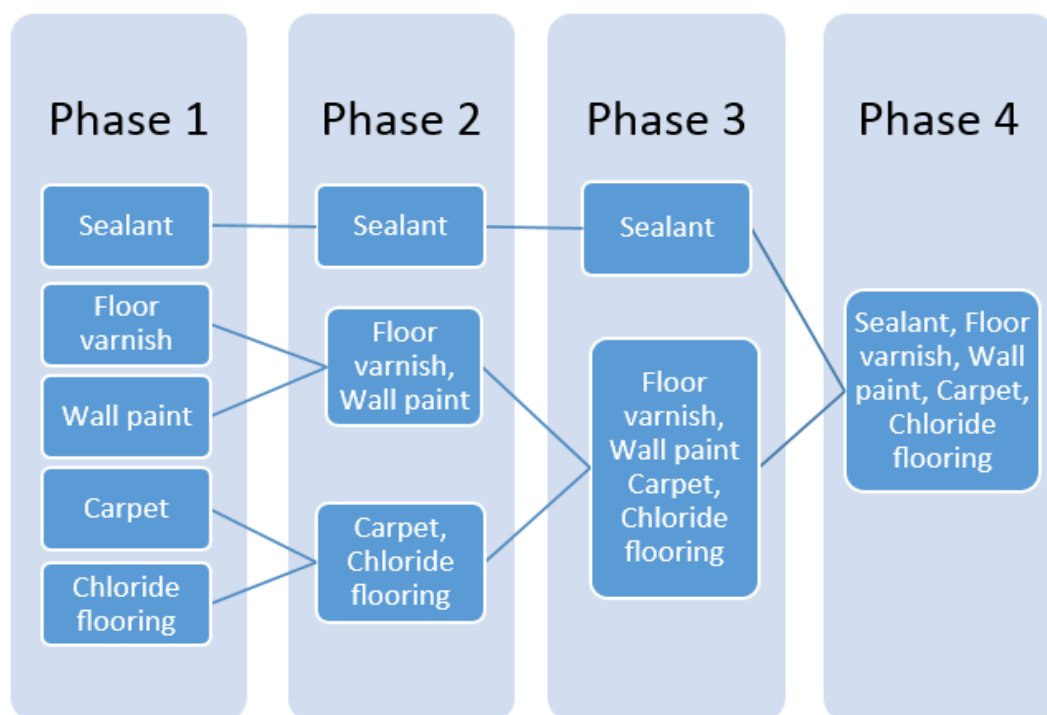
Giai đoạn	CK1	HPC1	CK2, HPC2, HPC3, HPC4
2	0.833	0.667	1
3	0.875	0.762	1

Bảng 3.11. Bảng chỉ số Purity của CK1, CK2, HPC1, HPC2, HPC3, HPC4 trên bộ dữ liệu xe hơi Quảng Châu

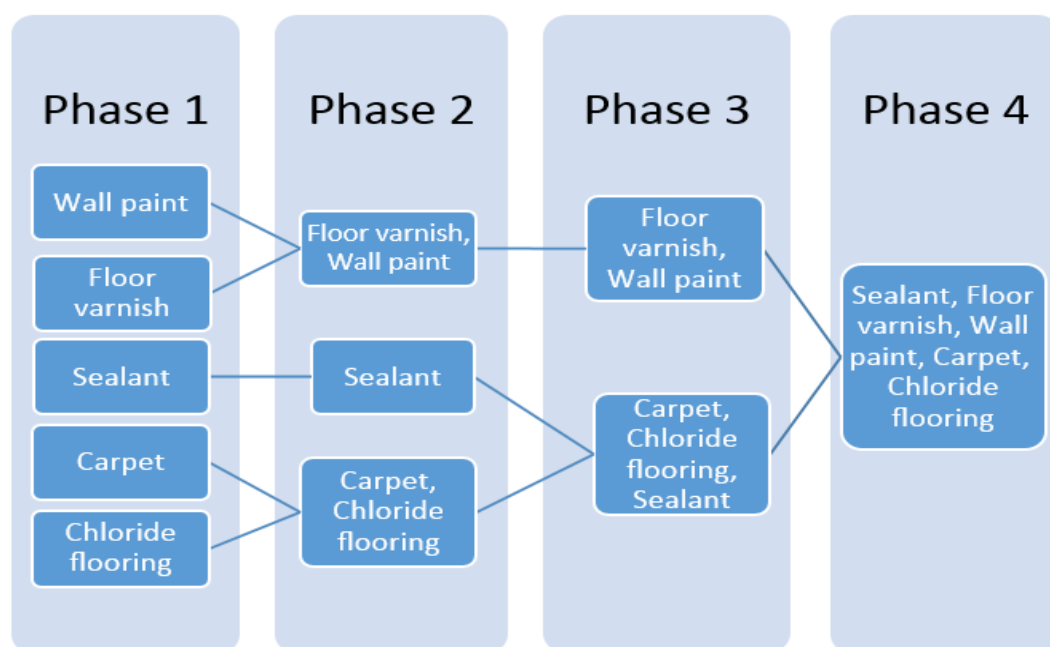
Giai đoạn	CK1	HPC1	CK2, HPC2, HPC3, HPC4
2	0.8	0.6	1
3	0.8	0.8	1

Kết quả thử nghiệm trên tập dữ liệu xe hơi Quảng Châu cho thấy rằng các thuật toán đưa ra với các độ đo khoảng cách viển cảnh tổng quát là tốt hơn của Cuong & Kreinovich. Điều này chứng tỏ rằng việc sử dụng các độ đo khoảng cách viển cảnh tổng quát trong các thuật toán phân cụm mang lại kết quả chất lượng phân cụm tốt hơn việc sử dụng các độ đo khoảng cách viển cảnh của Cuong & Kreinovich.

Để kiểm chứng tính hiệu quả của các thuật toán đưa ra, một thực nghiệm khác trên bộ dữ liệu khác đã được thực hiện đó là bộ dữ liệu vật liệu xây dựng trong bảng 3.2. Hình 3.8 minh họa cây phân cấp của thuật toán CK2 và HPC4, còn hình 3.9 là cây phân cấp của các thuật toán còn lại.



Hình 3.8: Cây phân cấp của thuật toán CK2 & HPC4 trên bộ dữ liệu vật liệu xây dựng



Hình 3.9: Cây phân cấp của IHC, CK1, HPC1, HPC2, HPC3 trên vật liệu xây dựng

Kết quả trong hình 3.9 khẳng định rõ ràng sự ổn định và không bị ảnh hưởng bởi các giá trị ngoại lệ của các thuật toán HPC đưa ra. Trong ví dụ, vật liệu “Sealant” tạo thành một cụm duy nhất từ giai đoạn 1 đến giai đoạn 3 như hình 3.8, ngược lại, cụm “Sealant” không xuất hiện trong giai đoạn 3 của hình 3.9. Các thuật toán HPC có xu hướng kết hợp các cụm đơn vào thành các cụm đặc nên các giá trị ngoại lệ có thể

được loại bỏ. Với kết quả phân cụm như ở bảng 3.12 thực hiện trên bộ dữ liệu vật liệu xây dựng ở bảng 3.2, ta thu được các bảng kết quả chỉ số đánh giá NMI, F-Measure, Purity từ bảng 3.13 – 3.15 như dưới đây.

Bảng 3.12. Bảng kết quả phân cụm của IHC, CK1, CK2, HPC1, HPC2, HPC3, HPC4 trên bộ dữ liệu vật liệu xây dựng

Giai đoạn	IHC	CK2, HPC4	CK1,HPC1,HPC2,HPC3
2	Floor varnish,	Floor varnish,	Floor varnish,
	Wall paint	Wall paint	Wall paint
	Carpet,	Carpet,	Carpet,
	Chloride flooring	Chloride flooring	Chloride flooring
3	Sealant	Sealant	Sealant
	Floor varnish,	Floor varnish,	Floor varnish,
	Wall paint	Wall paint, Carpet,	Wall paint
		Chloride flooring	
3	Carpet,	Sealant	Carpet,
	Chloride flooring,		Chloride flooring,
	Sealant		Sealant

Bảng 3.13. Bảng chỉ số NMI của CK1, CK2, HPC1, HPC2, HPC3, HPC4 trên bộ dữ liệu vật liệu xây dựng

Giai đoạn	CK2, HPC4	CK1, HPC1, HPC2, HPC3
2	1	1
3	0.204	1

Bảng 3.14. Bảng chỉ số F-Measure của CK1, CK2, HPC1, HPC2, HPC3, HPC4 trên bộ dữ liệu vật liệu xây dựng

Giai đoạn	CK2, HPC4	CK1, HPC1, HPC2, HPC3
2	1	1
3	0.952	1

Bảng 3.15. Bảng chỉ số Purity của CK1, CK2, HPC1, HPC2, HPC3, HPC4 trên bộ dữ liệu vật liệu xây dựng

Giai đoạn	CK2, HPC4	CK1, HPC1, HPC2, HPC3
2	1	1
3	0.8	1

Thực nghiệm trên bộ dữ liệu về bệnh tim, ta thu được kết quả bảng chỉ số đánh giá NMI, F-Measure, Purity từ bảng 3.16 – 3.18 như dưới đây.

Bảng 3.16. Bảng chỉ số NMI của CK1, CK2, HPC1, HPC2, HPC3, HPC4 trên bộ dữ liệu bệnh tim

Giai đoạn	CK1	CK2	HPC1	HPC2	HPC3	HPC4
2	0.882	0.88	0.882	0.881	0.881	0.882
3	0.749	0.74	0.743	0.744	0.751	0.748
4	0.612	0.609	0.612	0.601	0.628	0.604
5	0.474	0.43	0.479	0.465	0.52	0.491
6	0.321	0.353	0.371	0.353	0.393	0.364
7	0.216	0.265	0.224	0.296	0.273	0.219
8	0.109	0.164	0.122	0.138	0.113	0.116
9	0.004	0.027	0	0.024	0.027	0.027

Bảng 3.17. Bảng chỉ số F-Measure của CK1, CK2, HPC1, HPC2, HPC3, HPC4 trên bộ dữ liệu bệnh tim

Giai đoạn	CK1	CK2	HPC1	HPC2	HPC3	HPC4
2	0.581	0.574	0.581	0.578	0.578	0.581
3	0.434	0.408	0.426	0.423	0.441	0.423
4	0.4	0.397	0.388	0.374	0.415	0.382
5	0.379	0.327	0.389	0.364	0.432	0.414
6	0.367	0.397	0.428	0.395	0.456	0.452
7	0.472	0.452	0.422	0.471	0.462	0.436
8	0.514	0.573	0.522	0.417	0.467	0.48
9	0.637	0.785	0.747	0.609	0.785	0.785

Bảng 3.18. Bảng chỉ số Purity của CK1, CK2, HPC1, HPC2, HPC3, HPC4 trên bộ dữ liệu bệnh tim

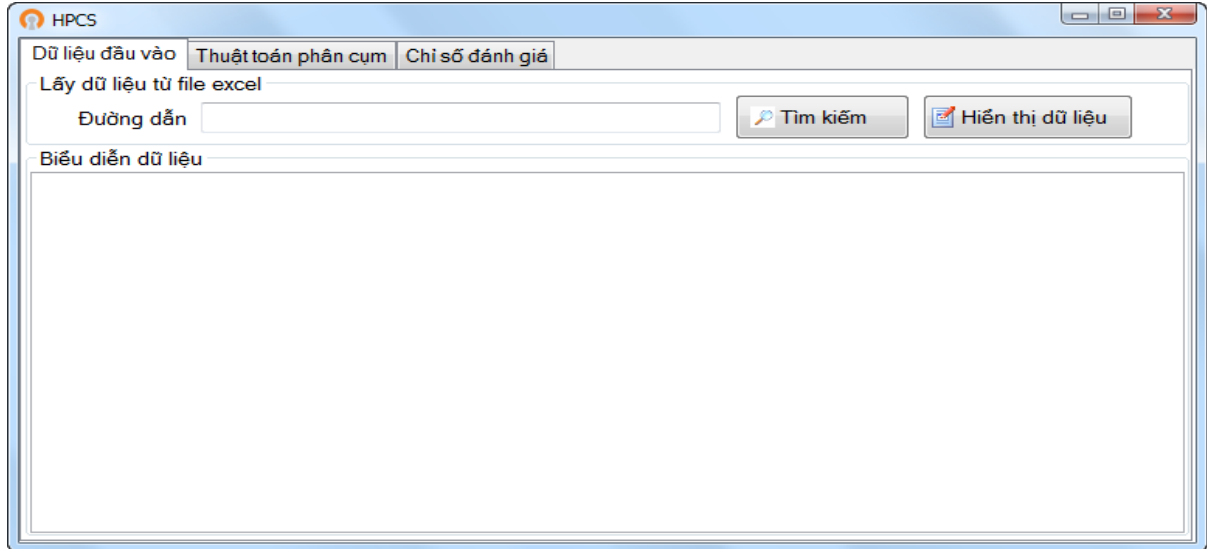
Giai đoạn	CK1	CK2	HPC1	HPC2	HPC3	HPC4
2	0.581	0.574	0.581	0.578	0.578	0.581
3	0.441	0.415	0.43	0.422	0.444	0.426
4	0.411	0.396	0.415	0.381	0.437	0.396
5	0.4	0.337	0.396	0.385	0.43	0.4
6	0.385	0.396	0.43	0.433	0.452	0.441
7	0.463	0.459	0.485	0.504	0.53	0.481
8	0.648	0.6	0.593	0.622	0.593	0.593
9	0.941	0.941	0.941	0.941	0.941	0.941

Từ các kết quả về bảng chỉ số NMI, Purity và F-Measure trên các bộ dữ liệu trên chúng ta có thể thấy số lượng các biến thể của HPC đạt chỉ số NMI, Purity và F-Measure cao thường nhiều hơn của CK. Do vậy các biến thể của HPC cho chất lượng tốt hơn của CK.

3.3.4. Hệ thống HPCS

Mục đích xây dựng hệ thống HPCS để thực nghiệm thuật toán HPC nhằm thực hiện phân cụm trên các bộ dữ liệu mờ viễn cảnh PFS khác nhau. Đồng thời cung cấp một công cụ hỗ trợ trực quan hóa phân cụm trên các bộ dữ liệu PFS.

Hệ thống HPCS, một hệ thống dựa trên tri thức, được xây dựng hoàn toàn mới bằng ngôn ngữ lập trình C# trên bộ công cụ lập trình Visual Studio 2010, chạy trên nền tảng NetFramework 3.5. Hệ thống HPCS có giao diện chính như hình 3.10.



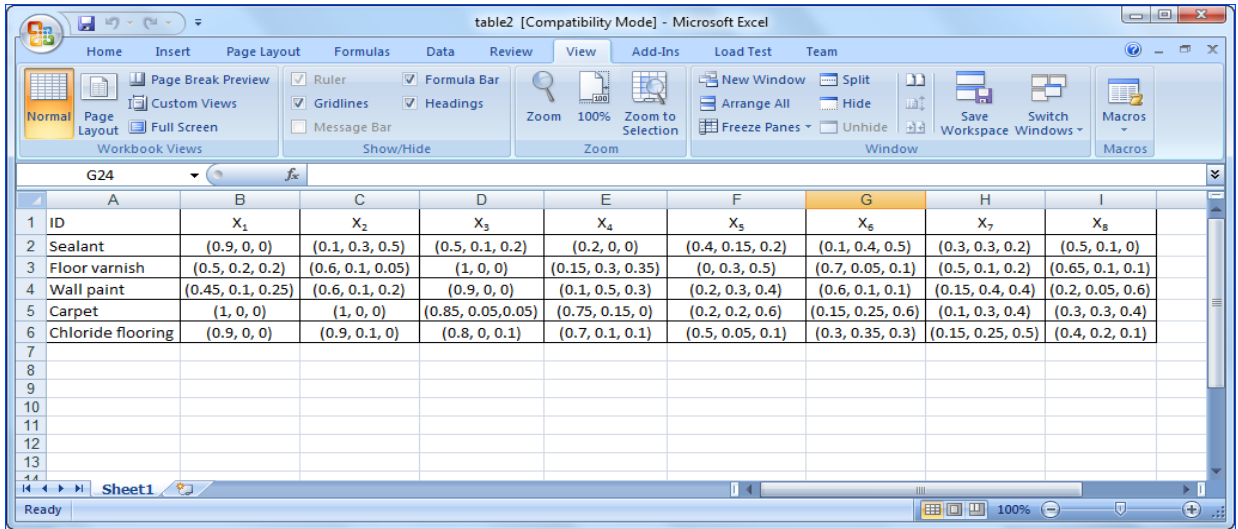
Hình 3.10: Giao diện chính của hệ thống HPCS

Hệ thống HPCS gồm hai phần chính là dữ liệu đầu vào và thuật toán phân cụm.

Dữ liệu đầu vào của hệ thống HPCS được lưu trong tệp Excel. Ví dụ như dữ liệu thực nghiệm ở bảng 3.1 – 3.3 được lưu trong tệp Excel tương ứng như hình 3.11 – 3.13. Hệ thống HPCS cho phép người dùng bấm nút “Tìm kiếm” để xác định đường dẫn đến tệp Excel chứa dữ liệu đầu vào. Sau đó, người dùng bấm nút “Hiển thị dữ liệu” để hệ thống đọc dữ liệu từ tệp Excel và biểu diễn dữ liệu dưới dạng bảng như hình 3.14.

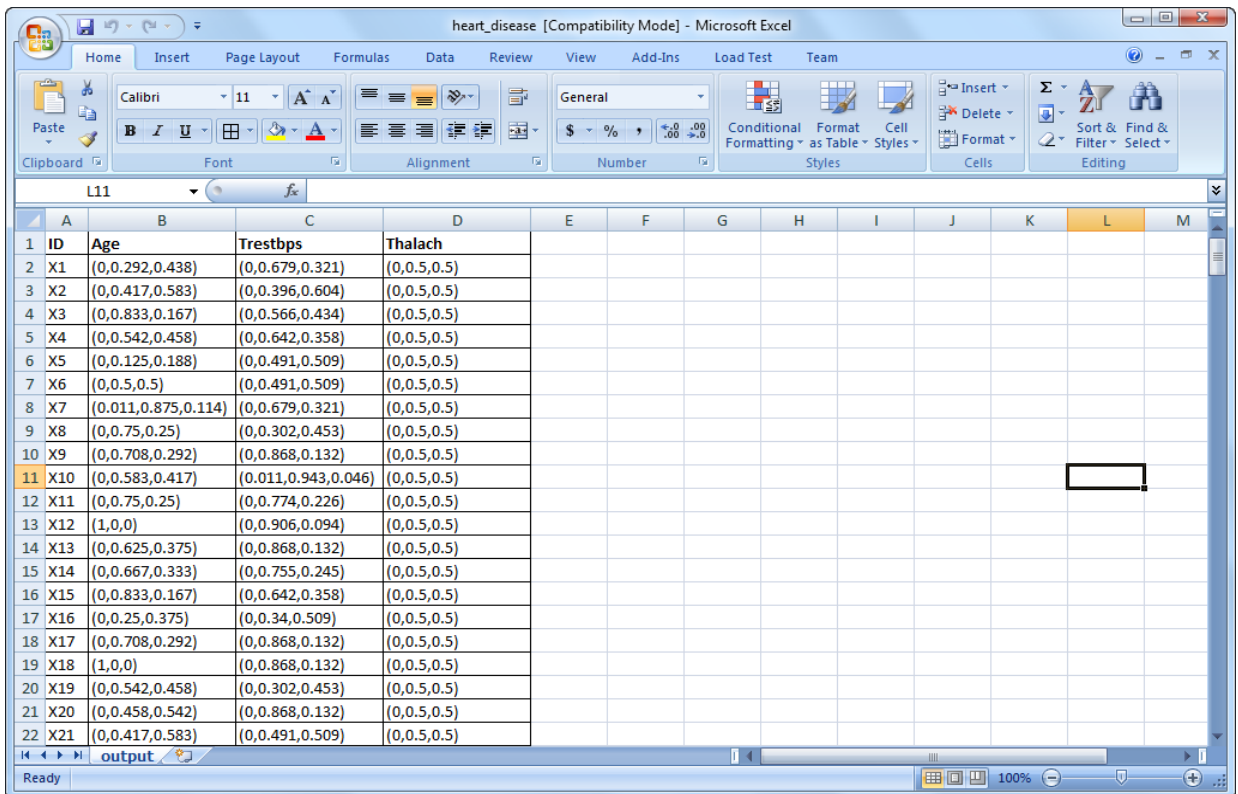
	A	B	C	D	E	F	G	H
	ID	Fuel(G ₁)	Aerod(G ₂)	Price(G ₃)	Comfort(G ₄)	Design(G ₅)	Safety(G ₆)	
1	Car1	(0.3, 0.1, 0.4)	(0.6, 0.05, 0.05)	(0.4, 0.1, 0.2)	(0.8, 0, 0.1)	(0.1, 0.2, 0.4)	(0.5, 0.3, 0.1)	
2	Car2	(0.6, 0.2, 0.1)	(0.5, 0.1, 0.1)	(0.6, 0, 0.1)	(0.1, 0.1, 0)	(0.3, 0.3, 0.3)	(0.4, 0.1, 0.2)	
3	Car3	(0.4, 0.3, 0.2)	(0.1, 0.5, 0.3)	(0.1, 0.2, 0.4)	(0.2, 0.1, 0.5)	(0.3, 0.4, 0.2)	(0.6, 0.3, 0.1)	
4	Car4	(0.2, 0.2, 0.2)	(0.4, 0.05, 0.05)	(0.9, 0, 0)	(0.8, 0.1, 0)	(0.2, 0.2, 0.3)	(0.7, 0.03, 0.07)	
5	Car5	(0, 0.4, 0.1)	(0.4, 0.1, 0.5)	(0.1, 0.6, 0.2)	(0.1, 0.1, 0)	(0.6, 0.1, 0.1)	(0.2, 0.1, 0.2)	
6								
7								
8								
9								
10								
11								
12								
13								
14								

Hình 3.11: Minh họa dữ liệu đầu vào ở bảng 3.1 lưu trong tệp Excel



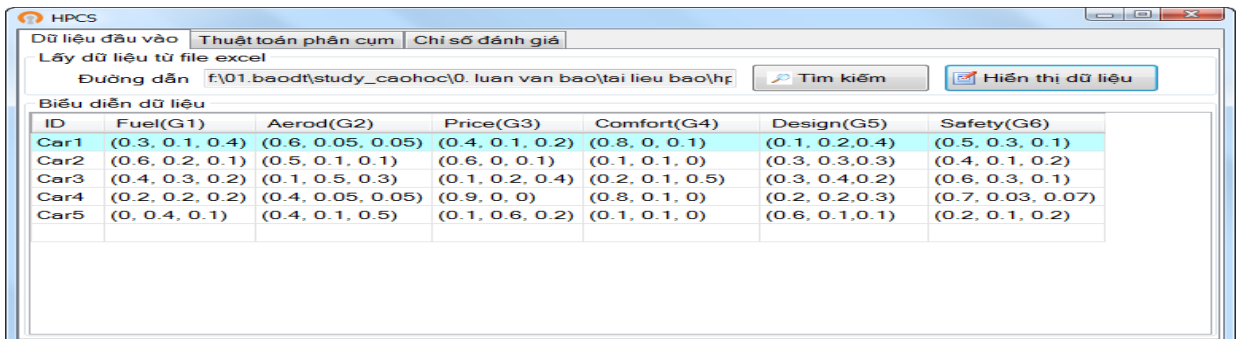
	A	B	C	D	E	F	G	H	I
1 ID		X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8
2 Sealant		(0.9, 0, 0)	(0.1, 0.3, 0.5)	(0.5, 0.1, 0.2)	(0.2, 0, 0)	(0.4, 0.15, 0.2)	(0.1, 0.4, 0.5)	(0.3, 0.3, 0.2)	(0.5, 0.1, 0)
3 Floor varnish		(0.5, 0.2, 0.2)	(0.6, 0.1, 0.05)	(1, 0, 0)	(0.15, 0.3, 0.35)	(0, 0.3, 0.5)	(0.7, 0.05, 0.1)	(0.5, 0.1, 0.2)	(0.65, 0.1, 0.1)
4 Wall paint		(0.45, 0.1, 0.25)	(0.6, 0.1, 0.2)	(0.9, 0, 0)	(0.1, 0.5, 0.3)	(0.2, 0.3, 0.4)	(0.6, 0.1, 0.1)	(0.15, 0.4, 0.4)	(0.2, 0.05, 0.6)
5 Carpet		(1, 0, 0)	(1, 0, 0)	(0.85, 0.05, 0.05)	(0.75, 0.15, 0)	(0.2, 0.2, 0.6)	(0.15, 0.25, 0.6)	(0.1, 0.3, 0.4)	(0.3, 0.3, 0.4)
6 Chloride flooring		(0.9, 0, 0)	(0.9, 0.1, 0)	(0.8, 0, 0.1)	(0.7, 0.1, 0.1)	(0.5, 0.05, 0.1)	(0.3, 0.35, 0.3)	(0.15, 0.25, 0.5)	(0.4, 0.2, 0.1)
7									
8									
9									
10									
11									
12									
13									

Hình 3.12: Minh họa dữ liệu đầu vào ở bảng 3.2 lưu trong tệp Excel



	A	B	C	D	E	F	G	H	I	J	K	L	M
1 ID		Age	Trestbps	Thalach									
2 X1		(0,0.292,0.438)	(0,0.679,0.321)	(0,0.5,0.5)									
3 X2		(0,0.417,0.583)	(0,0.396,0.604)	(0,0.5,0.5)									
4 X3		(0,0.833,0.167)	(0,0.566,0.434)	(0,0.5,0.5)									
5 X4		(0,0.542,0.458)	(0,0.642,0.358)	(0,0.5,0.5)									
6 X5		(0,0.125,0.188)	(0,0.491,0.509)	(0,0.5,0.5)									
7 X6		(0,0.5,0.5)	(0,0.491,0.509)	(0,0.5,0.5)									
8 X7		(0,0.11,0.875,0.114)	(0,0.679,0.321)	(0,0.5,0.5)									
9 X8		(0,0.75,0.25)	(0,0.302,0.453)	(0,0.5,0.5)									
10 X9		(0,0.708,0.292)	(0,0.868,0.132)	(0,0.5,0.5)									
11 X10		(0,0.583,0.417)	(0,0.11,0.943,0.046)	(0,0.5,0.5)									
12 X11		(0,0.75,0.25)	(0,0.774,0.226)	(0,0.5,0.5)									
13 X12		(1,0,0)	(0,0.906,0.094)	(0,0.5,0.5)									
14 X13		(0,0.625,0.375)	(0,0.868,0.132)	(0,0.5,0.5)									
15 X14		(0,0.667,0.333)	(0,0.755,0.245)	(0,0.5,0.5)									
16 X15		(0,0.833,0.167)	(0,0.642,0.358)	(0,0.5,0.5)									
17 X16		(0,0.25,0.375)	(0,0.34,0.509)	(0,0.5,0.5)									
18 X17		(0,0.708,0.292)	(0,0.868,0.132)	(0,0.5,0.5)									
19 X18		(1,0,0)	(0,0.868,0.132)	(0,0.5,0.5)									
20 X19		(0,0.542,0.458)	(0,0.302,0.453)	(0,0.5,0.5)									
21 X20		(0,0.458,0.542)	(0,0.868,0.132)	(0,0.5,0.5)									
22 X21		(0,0.417,0.583)	(0,0.491,0.509)	(0,0.5,0.5)									

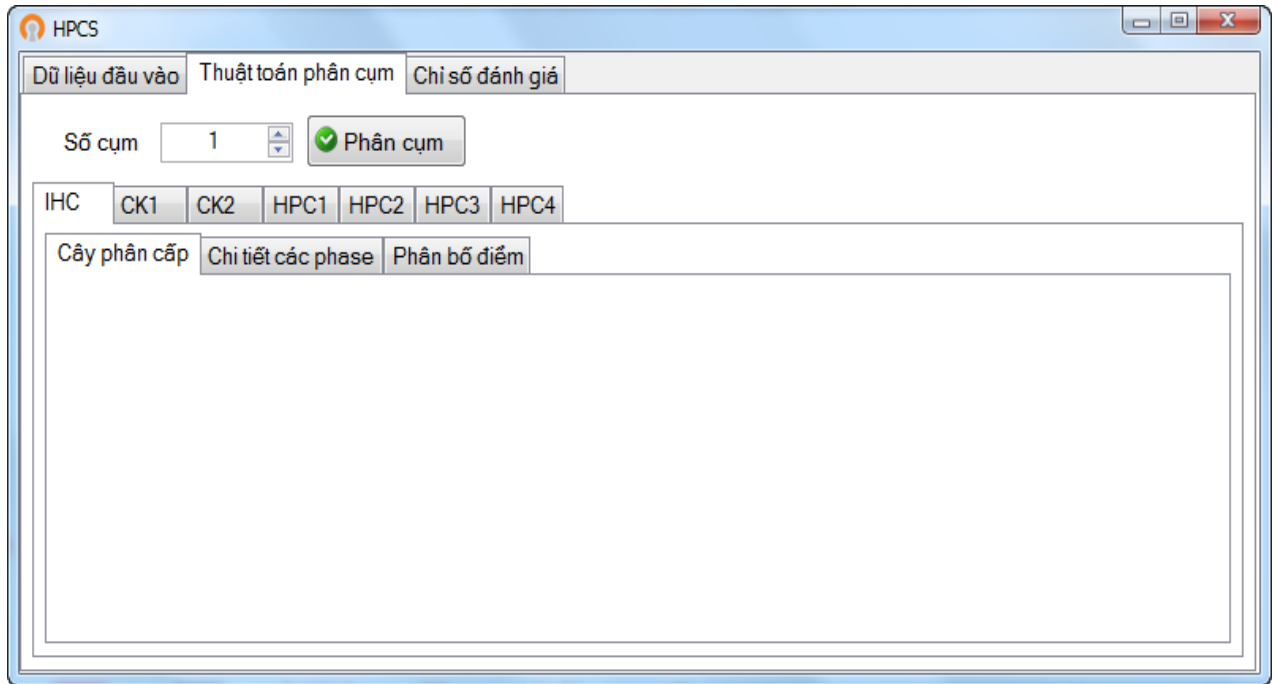
Hình 3.13: Minh họa dữ liệu đầu vào ở bảng 3.3 lưu trong tệp Excel



ID	Fuel(G1)	Aerod(G2)	Price(G3)	Comfort(G4)	Design(G5)	Safety(G6)
Car1	(0.3, 0.1, 0.4)	(0.6, 0.05, 0.05)	(0.4, 0.1, 0.2)	(0.8, 0, 0.1)	(0.1, 0.2, 0.4)	(0.5, 0.3, 0.1)
Car2	(0.6, 0.2, 0.1)	(0.5, 0.1, 0.1)	(0.6, 0, 0.1)	(0.1, 0.1, 0)	(0.3, 0.3, 0.3)	(0.4, 0.1, 0.2)
Car3	(0.4, 0.3, 0.2)	(0.1, 0.5, 0.3)	(0.1, 0.2, 0.4)	(0.2, 0.1, 0.5)	(0.3, 0.4, 0.2)	(0.6, 0.3, 0.1)
Car4	(0.2, 0.2, 0.2)	(0.4, 0.05, 0.05)	(0.9, 0, 0)	(0.8, 0.1, 0)	(0.2, 0.2, 0.3)	(0.7, 0.03, 0.07)
Car5	(0, 0.4, 0.1)	(0.4, 0.1, 0.5)	(0.1, 0.6, 0.2)	(0.1, 0.1, 0)	(0.6, 0.1, 0.1)	(0.2, 0.1, 0.2)

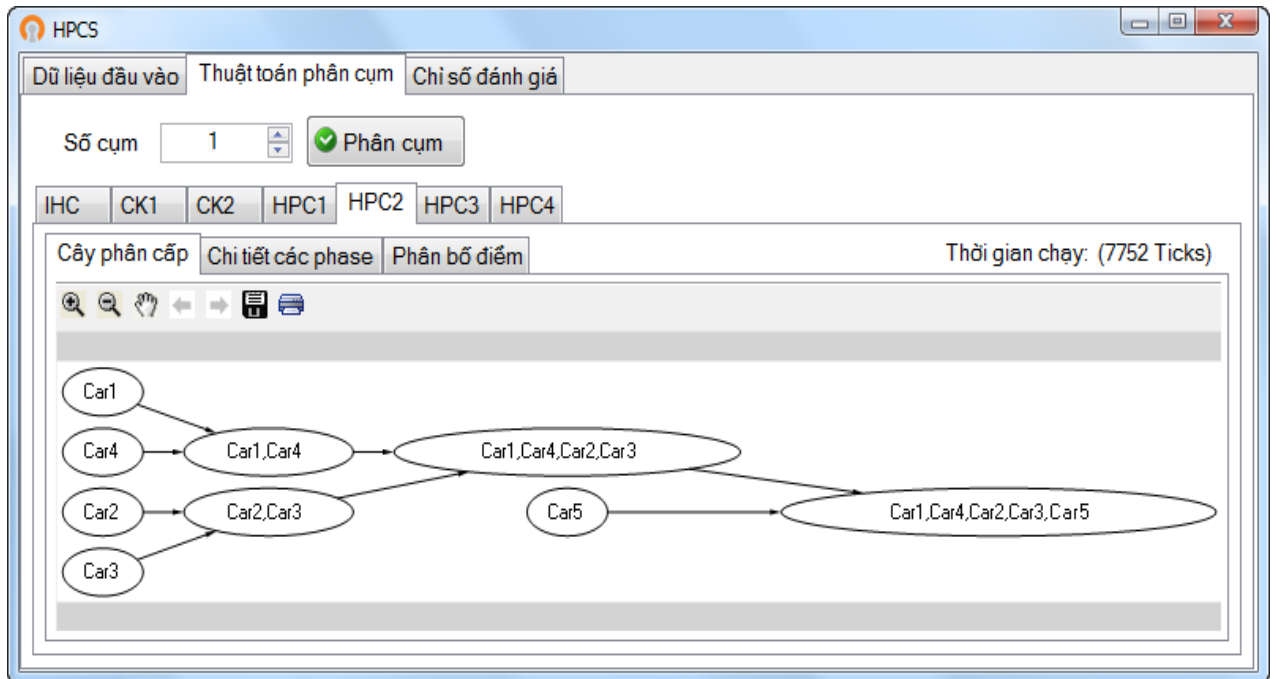
Hình 3.14: Biểu diễn dữ liệu đầu vào trên hệ thống HPSCS

Phần thuật toán phân cụm sẽ gồm 7 phiên bản IHC, CK1, CK2, HPC1, HPC2, HPC3, HPC4 như hình 3.15.

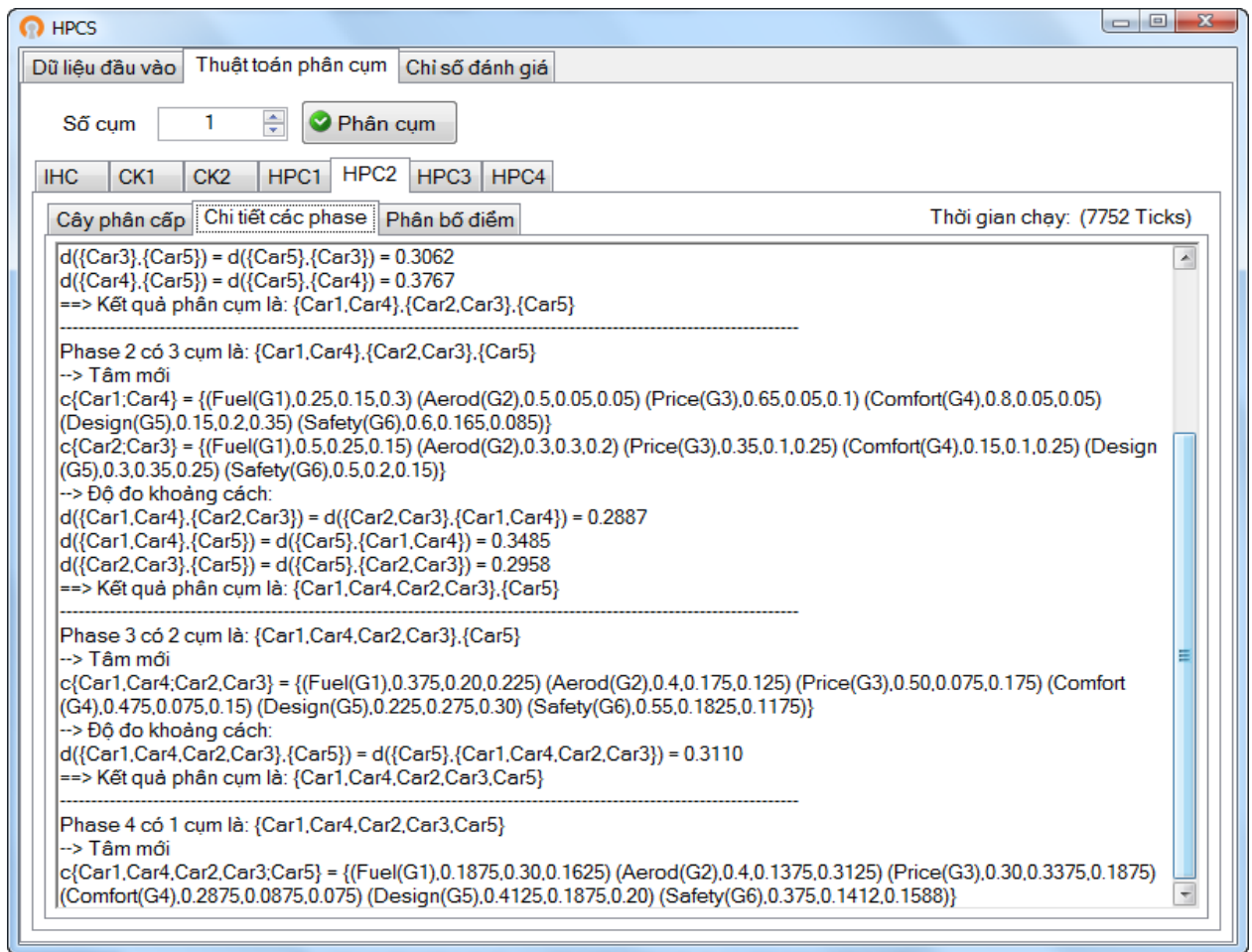


Hình 3.15: Các phiên bản thuật toán phân cụm trên hệ thống HPCS

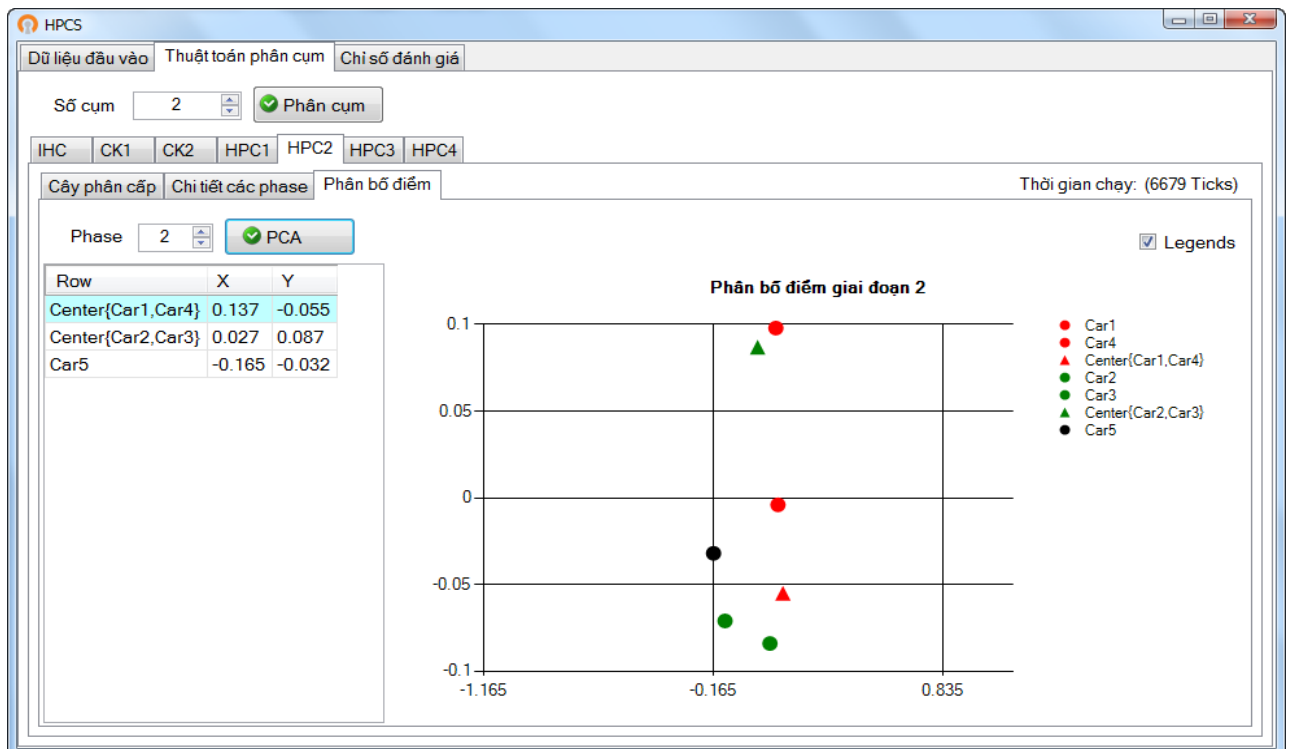
Sau khi bấm nút “Phân cụm” thì kết quả phân cụm tương ứng của từng thuật toán sẽ được biểu diễn dưới ba dạng là cây phân cấp như hình 3.16, số liệu tính toán qua từng giai đoạn của thuật toán như hình 3.17 và phân bố điểm và tâm (áp dụng PCA để giảm xuống 2 chiều) theo từng giai đoạn thuật toán như hình 3.18.



Hình 3.16: Minh họa cây phân cấp thuật toán HPC2 trên hệ thống HPCS

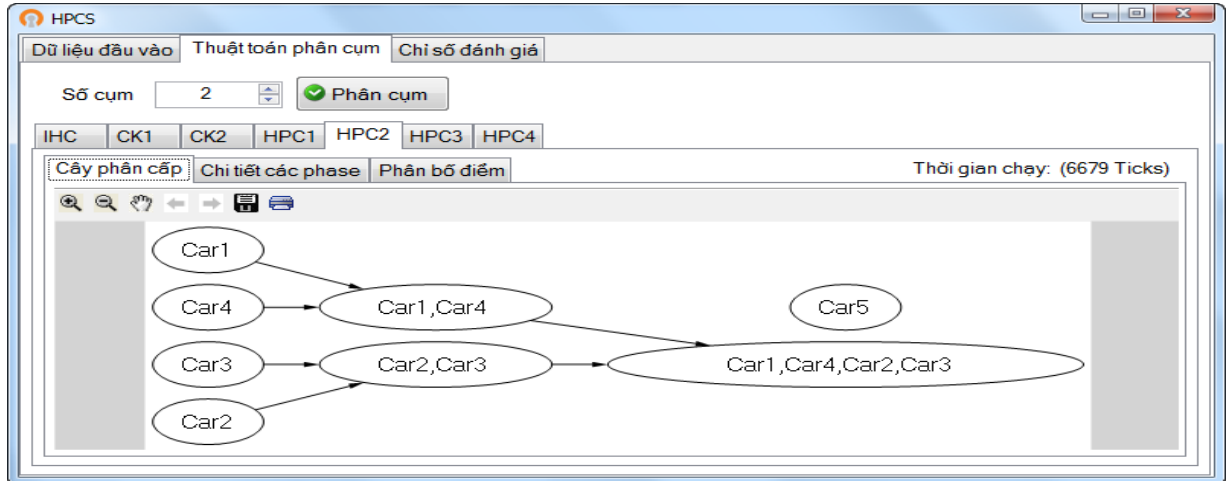


Hình 3.17: Minh họa số liệu tính toán qua các giai đoạn của thuật toán HPC2 trên hệ thống HPCS

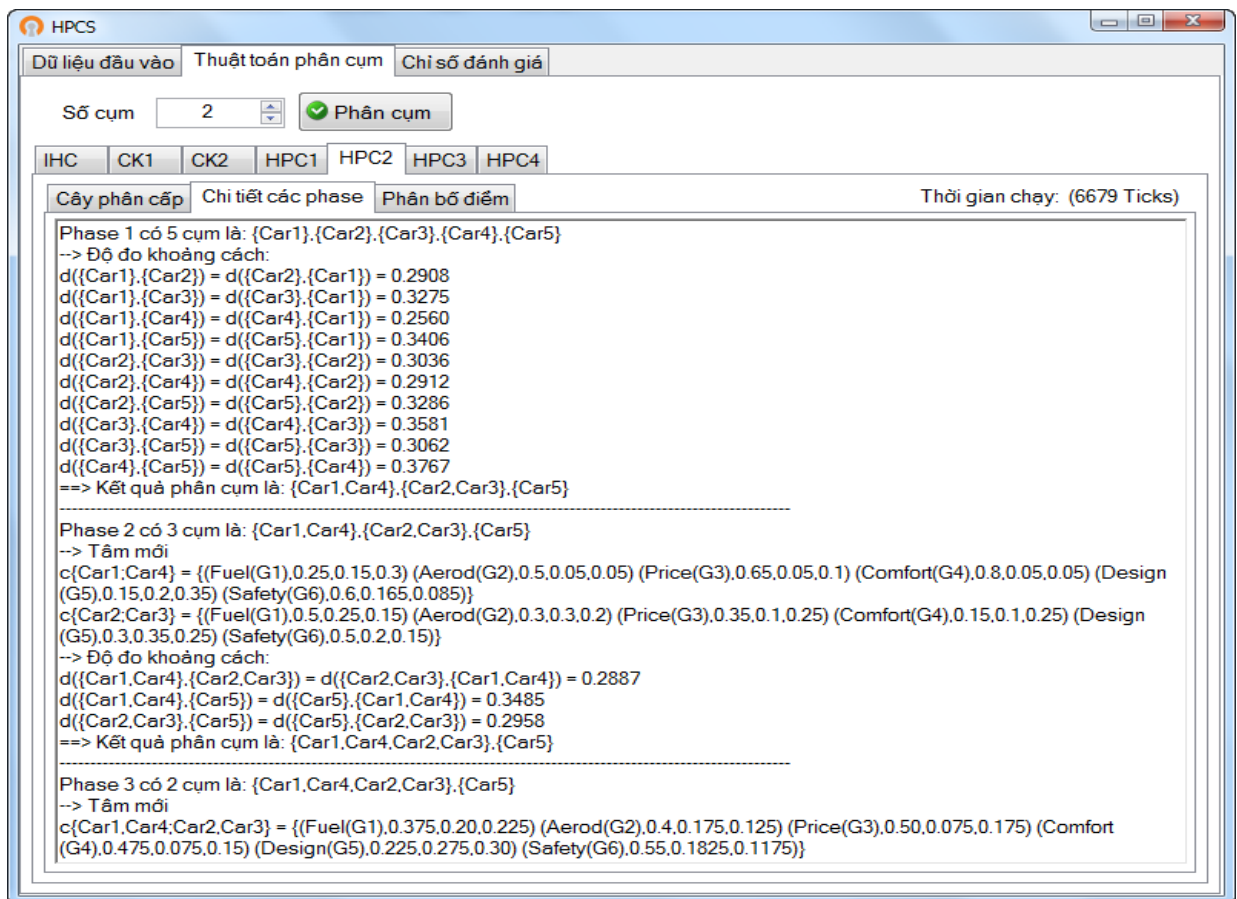


Hình 3.18: Minh họa phân bố điểm và tâm qua các giai đoạn của thuật toán HPC2 trên hệ thống HPCS

Bên cạnh đó, hệ thống HPCS còn cho phép xác định giá trị *số cụm mong muốn* để thuật toán phân cụm thực hiện lặp lại cho đến khi đạt được số cụm mong muốn bằng với giá trị tham số truyền vào. Giả sử số cụm mong muốn bằng 1 thì kết quả phân cụm qua các giai đoạn của thuật toán HPC2 sẽ như hình 3.16 – 3.17. Nhưng nếu số cụm là 2 thì lúc này kết quả phân cụm của thuật toán HPC2 sẽ như hình 3.19 – 3.20.



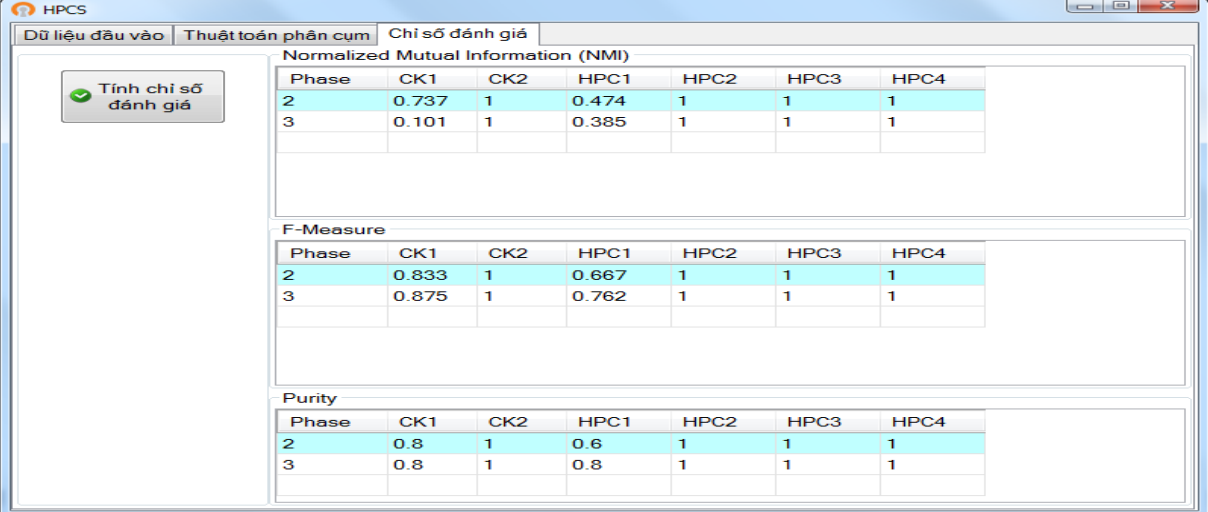
Hình 3.19: Minh họa cây phân cấp thuật toán HPC2 khi thay đổi số cụm trên hệ thống HPCS



Hình 3.20: Minh họa số liệu tính toán qua các giai đoạn của thuật toán HPC2 khi thay đổi số cụm trên hệ thống HPCS

Ngoài ra, trong thời gian thực hiện thuật toán, hệ thống HPCS sẽ đo và hiển thị thời gian chạy của từng phiên bản thuật toán phân cụm.

Đồng thời, hệ thống cho phép tính toán ra bảng chỉ số đánh giá NMI, F-Measure, Purity qua các giai đoạn của từng phiên bản thuật toán phân cụm so với mẫu chuẩn IHC. Sau khi đã thực hiện thuật toán phân cụm, thu được cây phân cấp, hệ thống cho phép bấm nút “Tính chỉ số đánh giá” để thực hiện tính toán và hiển thị bảng kết quả như hình 3.21 dưới đây.



Normalized Mutual Information (NMI)						
Phase	CK1	CK2	HPC1	HPC2	HPC3	HPC4
2	0.737	1	0.474	1	1	1
3	0.101	1	0.385	1	1	1

F-Measure						
Phase	CK1	CK2	HPC1	HPC2	HPC3	HPC4
2	0.833	1	0.667	1	1	1
3	0.875	1	0.762	1	1	1

Purity						
Phase	CK1	CK2	HPC1	HPC2	HPC3	HPC4
2	0.8	1	0.6	1	1	1
3	0.8	1	0.8	1	1	1

Hình 3.21: Minh họa bảng chỉ số đánh giá trên hệ thống HPCS

3.4. Kết luận chương

Chương 3 nghiên cứu về phân cụm trên dữ liệu mờ viễn cảnh PFS và đề xuất một số độ đo khoảng cách mờ viễn cảnh tổng quát được mở rộng từ độ đo viễn cảnh của Cuong & Kreinovich. Từ đó, sử dụng các độ đo này để thiết kế một phương pháp phân cụm phân cấp mờ viễn cảnh mới gọi là HPC. Kết quả thực nghiệm tiến hành trên các bộ dữ liệu khác nhau đã cho thấy rằng chất lượng phân cụm của các dạng thuật toán HPC là tốt hơn của Cuong & Kreinovich. Đồng thời, một hệ thống phân cụm có tên là HPCS đã được thiết kế để hỗ trợ trực quan hóa phân cụm trên dữ liệu mờ viễn cảnh PFS.

KẾT LUẬN

Sau khi lý thuyết tập mờ ra đời đã khẳng định được tính ưu việt bằng những kết quả thực tiễn đã mang lại và được rất nhiều nhà nghiên cứu quan tâm, phát triển. Một số mở rộng của tập mờ đã ra đời nhằm khắc phục những hạn chế của tập mờ truyền thống. Gần đây, một loại tập mờ tổng quát của các mở rộng nói trên đã được đề xuất đó là tập mờ viễn cảnh (hay còn gọi là tập mờ toàn cảnh) và hứa hẹn có khả năng ứng dụng cao. Chính vì vậy, với đề tài “Tập mờ viễn cảnh và ứng dụng”, luận văn tập trung tìm hiểu, nghiên cứu và đạt được một số kết quả sau:

- Trình bày các kiến thức cơ bản về tập mờ bao gồm định nghĩa, các phép toán tập hợp trên tập mờ, một số mở rộng của tập mờ và ứng dụng của nó, từ đó làm cơ sở để tìm hiểu, nghiên cứu tập mờ viễn cảnh PFS.
- Trình bày một số nội dung chính của tập mờ viễn cảnh PFS của Cuong & Kreinovich bao gồm định nghĩa, quan hệ giữa tập mờ viễn cảnh PFS với các tập mờ khác, các phép toán trên tập mờ viễn cảnh và một số tính chất, khoảng cách giữa các tập mờ viễn cảnh, quan hệ mờ viễn cảnh, v.v.
- Độ đo khoảng cách đóng vai trò quan trọng trong phân cụm và ảnh hưởng tới chất lượng phân cụm dữ liệu. Do đó, luận văn đã đề xuất một số độ đo khoảng cách viễn cảnh tổng quát và sử dụng chúng để thiết kế thuật toán phân cụm phân cấp mờ viễn cảnh mới gọi là HPC trên tập dữ liệu mờ viễn cảnh PFS.
- Cài đặt thực nghiệm thuật toán phân cụm phân cấp mờ viễn cảnh HPC với các độ đo viễn cảnh tổng quát đề xuất, các độ đo của Cuong & Kreinovich và thuật toán IHC [21]. Qua kết quả thực nghiệm khi tiến hành trên các bộ dữ liệu khác nhau với các chỉ số đánh giá đã cho thấy chất lượng phân cụm của HPC là tốt hơn của Cuong & Kreinovich. Đồng thời luận văn cung cấp một công cụ trực quan hóa phân cụm trên bộ dữ liệu mờ viễn cảnh PFS.

Hướng phát triển tiếp theo của đề tài sẽ là nghiên cứu một số tính chất toán học của các độ đo khoảng cách viễn cảnh tổng quát và ứng dụng trong phân cụm mờ.

TÀI LIỆU THAM KHẢO

1. Atanassov, K.T., Intuitionistic fuzzy sets, *Fuzzy Sets and Systems* 20 (1986) pp.87–96.
2. Chen, T.Y., A note on distances between intuitionistic fuzzy sets and/or interval-valued fuzzy sets based on the Hausdorff metric, *Fuzzy Sets Systems*.158(22) (2007), pp.2523–2525.
3. Cuong, B.C., Kreinovich, V., Picture Fuzzy Sets - a new concept for computational intelligence problems, *Proceeding of 2013 Third World Congress on Information and Communication Technologies (WICT 2013)*, pp.1-6.
4. Diamond, P., Kloeden, P., *Metric Spaces of Fuzzy Sets Theory and Applications*, World Scientific Publishing, Singapore, 1994
5. Dengfeng, L., Chuntian, C., New similarity measures of intuitionistic fuzzy sets and application to pattern recognitions, *Pattern Recognition Letters* 23(1) (2002) pp.221-225
6. Grzegorzewski, P., Distances between intuitionistic fuzzy sets and/or interval-valued fuzzy sets based on the Hausdorff metric, *Fuzzy Sets and Systems* 148(2) (2004), pp.319-328
7. Hatzimichailidis, A.G., Papakostas, G.A., Kaburlasos, V.G., A novel distance measure of intuitionistic fuzzy sets and its application to pattern recognition applications, *Int. J. Intell. Syst.* 27(4) (2012), pp.396–409
8. Hung, W. L., Yang, M. S., Similarity measures of intuitionistic fuzzy sets based on Hausdorff distance, *Pattern Recognition Letters* 25(14) (2004), pp. 1603-1611.
9. Hung, W. L., Yang, M. S., Similarity measures of intuitionistic fuzzy sets based on L_p metric, *International Journal of Approximate Reasoning* 46(1) (2007), pp.120-136
10. Liang, Z., Shi, P., Similarity measures on intuitionistic fuzzy sets, *Pattern Recognition Letters* 24(15) (2003), pp.2687-2693.
11. Li, Y., Olson, D. L., Qin, Z., Similarity measures between intuitionistic fuzzy (vague) sets: A comparative analysis, *Pattern Recognition Letters* 28(2) (2007), pp.278-285
12. Mitchell, H.B., On the Dengfeng–Chuntian similarity measure and its application to pattern recognition, *Pattern Recognition Letters* 24(16) (2003), pp.3101-3104
13. Papakostas, G. A., Hatzimichailidis, A. G., Kaburlasos, V. G., Distance and similarity measures between intuitionistic fuzzy sets: A comparative analysis from a pattern recognition point of view, *Pattern Recognition Letters* 34(14) (2013), pp.1609-1622
14. Smith, L.I, *A tutorial on Principal Components Analysis* (2002)
15. Szmiedt, E., Kacprzyk, J., Distances between intuitionistic fuzzy sets, *Fuzzy sets and systems* 114(3) (2000), pp.505-518.

16. Szmidt, E., Kacprzyk, J., Distances between intuitionistic fuzzy sets and their applications in reasoning, Springer Berlin Heidelberg, US, 2005.
17. Szmidt, E., Kacprzyk, J., Distances between intuitionistic fuzzy sets: straightforward approaches may not work, Proceedings of 3rd International IEEE Conference on Intelligent Systems (2006), pp.716-721
18. UCI Machine Learning Repository (2013). Heart Disease. URL: <https://archive.ics.uci.edu/ml/datasets/Heart+Disease>
19. Wang, W., Xin, X., Distance measure between intuitionistic fuzzy sets, Pattern Recognition Letters 26(13) (2005), pp.2063-2069
20. Xu, Z., Some similarity measures of intuitionistic fuzzy sets and their applications to multiple attribute decision making, Fuzzy Optimization and Decision Making 6(2) (2007), pp.109-121.
21. Xu, Z., Chen, J., Wu, J., Clustering algorithm for intuitionistic fuzzy sets, Information Sciences 178(19) (2008), pp.3775-3790.
22. Xu, Z. S., Chen, J., An overview of distance and similarity measures of intuitionistic fuzzy sets, Int. J. Uncertain Fuzziness Knowl-Based Syst. 16 (2008), pp.529–555.
23. Xu, Z. S, Intuitionistic fuzzy hierarchical clustering algorithms, Journal of Systems Engineering and Electronics 20 (2009), pp.90–97.
24. Xu, Z., Choquet integrals of weighted intuitionistic fuzzy information, Information Sciences 180(5) (2010), pp.726-736.
25. Xu, Z., Xia, M., Distance and similarity measures for hesitant fuzzy sets, Information Sciences 181(11) (2011), pp.2128-2138.
26. Xu, Z., Intuitionistic fuzzy aggregation and clustering, Springer, US, 2012, pp.192-199.
27. Yang, Y., Chiclana, F., Consistency of 2D and 3D distances of intuitionistic fuzzy sets, Expert Systems with Applications (2012).
28. Zadeh, L.A., Fuzzy Sets, Information and Control (1965) pp.338–353.
29. Zadeh, L.A., “The Concept of a Linguistic Variable and Its Application to Approximate Reasoning–1,” Information Sciences 8 (1975), pp.199–249.
30. Zhang, H. M., Xu, Z. S., Chen, Q., On clustering approach to intuitionistic fuzzy sets, Control Decision 22 (2007), pp.882–888.