

**BỘ GIÁO DỤC
VÀ ĐÀO TẠO**

**VIỆN HÀN LÂM KHOA HỌC
VÀ CÔNG NGHỆ VIỆT NAM**

HỌC VIỆN KHOA HỌC VÀ CÔNG NGHỆ



Nguyễn Thanh Sơn

**NGHIÊN CỨU PHÁT TRIỂN MỘT SỐ MÔ HÌNH PHÂN CỤM BÁN GIÁM SÁT
DỰA TRÊN MẠNG NƠRON MIN-MAX MỜ**

TÓM TẮT LUẬN ÁN TIẾN SĨ MÁY TÍNH

Ngành: Hệ thống thông tin

Mã số: 9 48 01 04

Hà Nội- 2026

Công trình được hoàn thành tại: Học viện Khoa học và Công nghệ, Viện Hàn lâm Khoa học và Công nghệ Việt Nam

Người hướng dẫn khoa học:

Người hướng dẫn 1: PGS.TS. Lê Bá Dũng, Viện Công nghệ thông tin, Viện Hàn lâm KHCN VN

Người hướng dẫn 2: TS. Vũ Đình Minh, Trường Công nghệ thông tin và Truyền thông, Đại học Công nghiệp Hà Nội

Phản biện 1: PGS.TS. Nguyễn Văn Núi, Trường Đại học Công nghệ thông tin và Truyền thông, Đại học Thái Nguyên

Phản biện 2: PGS.TS. Phạm Văn Hải, Trường Đại học Công nghệ thông tin và Truyền thông, Đại học Bách khoa Hà Nội.

Phản biện 3: PGS.TS. Trần Thị Ngân, Trường Quốc tế, Đại học Quốc gia HN

Luận án được bảo vệ trước Hội đồng đánh giá luận án tiến sĩ cấp Học viện họp tại Học viện Khoa học và Công nghệ, Viện Hàn lâm Khoa học và Công nghệ Việt Nam vào hồi 14 giờ 00, ngày 08 tháng 9 năm 2026

Có thể tìm hiểu luận án tại:

1. Thư viện Học viện Khoa học và Công nghệ
2. Thư viện Quốc gia Việt Nam

MỞ ĐẦU

1. Tính cấp thiết của đề tài

Sự gia tăng nhanh chóng của dữ liệu về quy mô, số chiều và mức độ phức tạp đặt ra yêu cầu phát triển các phương pháp khai phá dữ liệu có khả năng học từ lượng thông tin giám sát hạn chế. Trong nhiều bài toán thực tế, dữ liệu chưa gán nhãn có thể thu thập với số lượng lớn, trong khi quá trình gán nhãn thường tốn thời gian, chi phí và cần kiến thức chuyên gia. Phân cụm bán giám sát vì vậy trở thành hướng tiếp cận phù hợp khi kết hợp cấu trúc nội tại của dữ liệu chưa gán nhãn với một lượng nhỏ thông tin giám sát. [1, 2]

Mạng nơron Min-Max mờ (FMN) có ưu điểm học gia tăng, biểu diễn dữ liệu bằng các siêu hộp mờ và không yêu cầu xác định trước số lượng cụm. Tuy nhiên, các mô hình FMN bán giám sát vẫn gặp khó khăn khi dữ liệu phân bố không đồng đều, chồng lấn hoặc nhiễu; đặc biệt, suy luận và lan truyền nhãn có thể phụ thuộc quá nhiều vào một siêu hộp hoặc một mẫu đại diện. Nếu nguồn thông tin này có độ tin cậy thấp, lỗi nhãn có thể được khuếch đại trong quá trình học. [15, 16, 17, 18, 19, 21]

Từ các hạn chế đó, luận án tập trung phát triển các mô hình phân cụm bán giám sát dựa trên FMN theo hai hướng: nâng cao độ tin cậy của suy luận nhãn bằng cơ chế bỏ phiếu đa số và cải thiện chất lượng tập hạt giống bằng chiến lược lựa chọn chủ động. Các mô hình FMN-Voting, FMN-SAL và FA-Seed được xây dựng để cụ thể hóa các hướng nghiên cứu này.

2. Mục tiêu nghiên cứu

Mục tiêu tổng quát của luận án là nghiên cứu, phát triển các mô hình phân cụm bán giám sát dựa trên mạng nơron Min-Max mờ nhằm nâng cao độ ổn định, độ tin cậy và hiệu năng học trong điều kiện thông tin nhãn hạn chế.

Các mục tiêu cụ thể tập trung vào: (i) nghiên cứu cơ chế khai thác cấu trúc siêu hộp và thông tin lân cận để suy luận nhãn tin cậy hơn; (ii) phát triển cơ chế bỏ phiếu đa số dựa trên siêu hộp và dựa trên các mẫu có nhãn lân cận; (iii) xây dựng chiến lược lựa chọn hạt giống chủ động, có xét đến mức độ đại diện và độ tin cậy; và (iv) thực nghiệm, so sánh các mô hình đề xuất trên dữ liệu chuẩn và dữ liệu thực tế.

3. Đối tượng và phạm vi nghiên cứu

Đối tượng nghiên cứu là các mô hình phân cụm bán giám sát dựa trên FMN, tập trung vào biểu diễn dữ liệu bằng siêu hộp, cơ chế suy luận và lan truyền nhãn, bỏ phiếu đa số và lựa chọn hạt giống. Phạm vi thực nghiệm gồm các tập dữ liệu chuẩn có cấu trúc cụm khác nhau và dữ liệu thực tế về bệnh gan, với các chỉ số đánh giá phân cụm và phân loại phù hợp.

4. Phương pháp nghiên cứu

Luận án sử dụng phương pháp nghiên cứu lý thuyết kết hợp thực nghiệm. Trên cơ sở tổng quan FMN và phân cụm bán giám sát, các hạn chế được phân tích để xây dựng mô hình, công thức và thuật toán cải tiến. Các mô hình được phân tích độ phức tạp tính toán và đánh giá thực nghiệm thông qua các chỉ số Accuracy, Precision, Recall, Specificity, F1-Score, RI, ARI và NMI.

5. Những đóng góp mới của luận án

Thứ nhất, luận án đề xuất mô hình FMN-Voting, sử dụng cơ chế bỏ phiếu đa số từ các siêu hộp lân cận nhằm giảm sự phụ thuộc vào một siêu hộp đơn lẻ trong suy luận nhãn.

Thứ hai, luận án đề xuất FMN-SAL, chuyển nguồn bỏ phiếu từ các siêu hộp sang các mẫu có nhãn lân cận, qua đó hạn chế ảnh hưởng của các siêu hộp có mức độ đại diện không đồng đều hoặc bị nhiễu.

Thứ ba, luận án đề xuất FA-Seed, một chiến lược lựa chọn hạt giống chủ động dựa trên FMN, xem xét đồng thời tính đại diện và độ tin cậy để nâng cao chất lượng tập hạt giống và hạn chế lan truyền lỗi nhãn.

6. Cấu trúc của luận án

Ngoài Mở đầu và Kết luận, luận án gồm ba chương.

Chương 1 trình bày tổng quan về phân cụm bán giám sát và FMN.

Chương 2 trình bày FMN-Voting và FMN-SAL.

Chương 3 trình bày mô hình FA-Seed và đánh giá thực nghiệm.

CHƯƠNG 1

TỔNG QUAN VỀ PHÂN CỤM BÁN GIÁM SÁT SỬ DỤNG MẠNG NƠON MIN-MAX MỜ

1.1. Tổng quan về bài toán phân cụm dữ liệu và ứng dụng

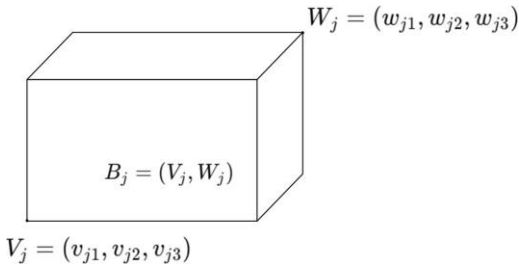
Phân cụm là kỹ thuật học không giám sát nhằm nhóm các đối tượng tương đồng vào cùng một cụm và tách các đối tượng khác biệt sang các cụm khác nhau. Trong phân cụm rõ, mỗi đối tượng thuộc duy nhất một cụm; trong phân cụm mờ, một đối tượng có thể thuộc nhiều cụm với các mức độ thuộc khác nhau. Phân cụm bán giám sát mờ kết hợp khả năng biểu diễn tính không chắc chắn của phân cụm mờ với một lượng nhỏ thông tin giám sát. [41, 42]

Các hướng tiếp cận tiêu biểu gồm phân cụm dựa trên mẫu hạt giống, ràng buộc cặp, mô hình nguyên mẫu và mạng nơon mờ. Các phương pháp dựa trên hạt giống sử dụng một tập nhỏ mẫu đã gán nhãn để định hướng quá trình phân cụm, giúp giảm phụ thuộc vào khởi tạo ngẫu nhiên và cải thiện độ ổn định. [43, 44, 45, 46, 47, 48]

1.2. Phân cụm bán giám sát dựa trên mạng nơon Min-Max mờ

1.2.1. Giới thiệu về mạng nơon Min-Max mờ

FMN biểu diễn dữ liệu bằng các siêu hộp mờ trong không gian đặc trưng. Mỗi siêu hộp B_j được xác định bởi hai vector đỉnh min V_j và đỉnh max W_j . Cấu trúc này cho phép mô hình học gia tăng, mở rộng siêu hộp để bao phủ mẫu mới và tạo siêu hộp mới khi không thể mở rộng các siêu hộp hiện có. [15, 16, 76]



Hình 1.3: Siêu hộp B_j trong không gian 3D với hai đỉnh V_j và W_j

1.2.2. Hàm thuộc siêu hộp mờ

Mức độ thuộc của mẫu x_r vào siêu hộp B_j , ký hiệu $\mu(x_r, B_j)$, phản ánh mức tương thích giữa mẫu và vùng không gian do siêu hộp biểu diễn. Tham số mờ γ điều khiển tốc độ suy giảm độ thuộc khi mẫu nằm ngoài biên siêu hộp.

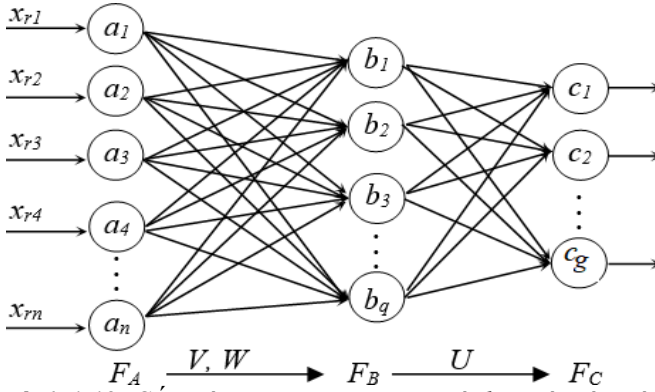
$$\mu(x_r, B_j) = \frac{1}{n} \sum_{i=1}^n \left[1 - f(x_{ri} - w_{ji}, \gamma) - f(v_{ji} - x_{ri}, \gamma) \right] \quad (1.3)$$

$$f(s, z) = \begin{cases} 1, & s \cdot z > 1, \\ s \cdot z, & 0 \leq s \cdot z \leq 1, \\ 0, & s \cdot z < 0. \end{cases} \quad (1.4)$$

Kích thước siêu hộp được kiểm soát bởi giới hạn δ^{max} . Khi một mẫu mới được xét, mô hình tìm siêu hộp phù hợp, kiểm tra điều kiện mở rộng, cập nhật các đỉnh và sau đó kiểm tra chồng lấn.

1.2.3. Cấu trúc mạng nơron FMN

Kiến trúc FMN gồm lớp đầu vào, lớp siêu hộp và lớp đầu ra. Lớp siêu hộp thực hiện biểu diễn các vùng dữ liệu và tính độ thuộc; lớp đầu ra tổng hợp thông tin từ các siêu hộp để xác định lớp hoặc cụm. Trong học có giám sát, liên kết giữa siêu hộp và lớp được biểu diễn bởi ma trận nhị phân.



Hình 1.12: Cấu trúc mạng nơron FMN với học có giám sát.

1.2.4. Thuật toán học FMN

Các thuật toán học FMN gồm học có giám sát, không giám sát và bán giám sát. Quá trình học dựa trên các thao tác: lựa chọn siêu hộp, kiểm tra điều kiện mở rộng, cập nhật đỉnh, kiểm tra chồng lấn và điều chỉnh chồng lấn. Trong học bán giám sát, dữ liệu có nhãn và chưa gán nhãn được khai thác đồng thời. [15, 16, 19]

Với mỗi mẫu, nếu tồn tại siêu hộp phù hợp và điều kiện kích thước được thỏa mãn thì siêu hộp được mở rộng. Nếu không thể mở rộng, một siêu hộp mới được tạo. Cơ chế học gia tăng giúp FMN thích nghi với dữ liệu mới mà không cần huấn luyện lại toàn bộ mạng.

Thuật toán 1.3 Thuật toán học bán giám sát trong FMN

Input: $X = \{(x_r, \ell_{x_r})\}_{r=1}^M, \gamma, \delta^{\max}, \beta$

Output: $\mathcal{B}$

```

1: repeat
2:   Chọn dữ liệu vào  $(x_r, \ell_{x_r}) \in X$ ;
3:   if  $\exists B_j \in \mathcal{B}$  thỏa mãn  $\mu(x_r, B_j) = \max_{B_k \in \mathcal{B}} \mu(x_r, B_k)$  và  $\delta(x_r, B_j) \leq \delta^{\max}$  then
4:     Điều chỉnh  $B_j$  theo (1.6) và (1.7);  $X \leftarrow X \setminus \{(x_r, \ell_{x_r})\}$ ;
5:     if phát hiện chồng lấn giữa  $B_j$  và  $B_k$  với  $\ell(B_j) \neq \ell(B_k)$  then
6:       Điều chỉnh chồng lấn giữa các siêu hộp  $B_j$  và  $B_k$ ;
7:     end if
8:   else
9:     if  $\ell(x_r) \neq 0$  then
10:      Tạo siêu hộp mới  $B_{\text{new}}$ ;  $\ell(B_{\text{new}}) \leftarrow \ell_{x_r}$ ;  $\mathcal{B} \leftarrow \mathcal{B} \cup \{B_{\text{new}}\}$ 
11:    else
12:      if  $\exists B_j \in \mathcal{B}$  sao cho  $\mu(x_r, g(B_j)) > \beta$  then
13:        Tạo siêu hộp mới  $B_{\text{new}}$ ;  $\ell(B_{\text{new}}) \leftarrow \ell(B_j)$ ;  $\mathcal{B} \leftarrow \mathcal{B} \cup \{B_{\text{new}}\}$ ;
14:      end if
15:       $\ell_{x_r} = \ell(B_j)$ ;
16:    end if
17:  until  $X = \emptyset$ 
18: return  $\mathcal{B}, \{(x_r, \ell_{x_r})\}_{r=1}^M$ 

```

1.2.5. Một số vấn đề cần tiếp tục nghiên cứu để nâng cao chất lượng FMN cho phân cụm dữ liệu

Chất lượng FMN phụ thuộc vào cấu trúc và kích thước siêu hộp. Khi $\delta^{\max}$ nhỏ, số lượng siêu hộp tăng; khi $\delta^{\max}$ lớn, chồng lấn có thể tăng và làm giảm khả năng phân biệt. Các siêu hộp cũng có mức độ đại diện và độ tin cậy khác nhau, nhưng nhiều mô hình vẫn xem chúng có vai trò tương đương trong lan truyền nhãn. [17, 18, 19]

Trong điều kiện nhãn hạn chế, việc lựa chọn hạt giống và nguồn suy luận nhãn trở nên đặc biệt quan trọng. Hạt giống ở vùng biên hoặc nhiễu có thể làm lan truyền sai nhãn; tương tự, một siêu hộp không ổn định có thể trở thành nguồn giám sát kém tin cậy.

1.3. Hạn chế, vấn đề đặt ra và định hướng nghiên cứu

Tổng quan cho thấy các mô hình FMN bán giám sát còn ba vấn đề chính. Thứ nhất, suy luận nhãn dựa trên một thực thể đơn lẻ có thể thiếu ổn định. Thứ hai, chất lượng của các siêu hộp lân cận không đồng đều, do đó bỏ phiếu trực tiếp từ siêu hộp vẫn có thể chịu ảnh hưởng của nhiễu và phân bố dữ liệu không đều. Thứ ba, lựa chọn ngẫu nhiên hoặc không thích nghi các hạt giống không bảo đảm tính đại diện trong điều kiện ngân sách nhãn hạn chế. [17, 18, 19, 21]

Từ đó, luận án định hướng: (i) xây dựng cơ chế bỏ phiếu đa số từ nhiều siêu hợp để tăng độ ổn định của suy luận nhãn; (ii) phát triển cơ chế bỏ phiếu dựa trên các mẫu có nhãn lân cận nhằm giảm ảnh hưởng của cấu trúc siêu hợp kém đại diện; và (iii) lựa chọn hạt giống chủ động dựa trên cấu trúc FMN, có xét đến tính đại diện và độ tin cậy.

1.4. Kết luận chương 1

Chương 1 đã hệ thống hóa cơ sở lý thuyết của phân cụm bán giám sát và FMN, làm rõ biểu diễn siêu hợp, hàm thuộc, cấu trúc mạng và cơ chế học. Các hạn chế được xác định là cơ sở trực tiếp cho FMN-Voting, FMN-SAL và FA-Seed ở các chương tiếp theo.

Một phần nội dung Chương 1 được công bố trong công trình [CT1].

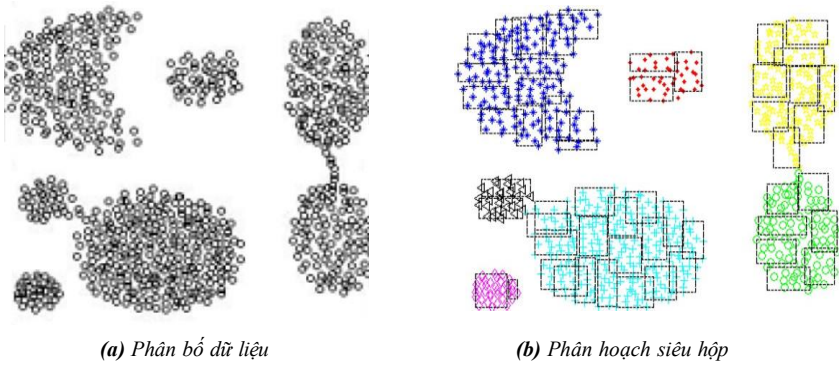
CHƯƠNG 2

PHÂN CỤM BÁN GIÁM SÁT DỰA TRÊN MẠNG NƠON MIN-MAX MỜ VỚI BỎ PHIẾU ĐA SỐ

2.1. Học bán giám sát dựa trên siêu hộp mờ

2.1.1. Phân hoạch không gian dữ liệu bằng FMN

FMN phân hoạch không gian dữ liệu thành các siêu hộp. Các siêu hộp vừa đóng vai trò đơn vị biểu diễn cấu trúc dữ liệu vừa là cơ sở để suy luận nhân. Cách biểu diễn này giảm số lượng đối tượng cần xử lý so với thao tác trực tiếp trên toàn bộ mẫu. [15, 16, 76]



Hình 2.1: Không gian siêu hộp được phân hoạch trên tập dữ liệu Aggregation

2.1.2. Đánh giá siêu hộp dựa trên mật độ và phân phối dữ liệu trong siêu hộp

Không phải mọi siêu hộp đều có chất lượng như nhau. Mật độ và phân phối các mẫu bên trong siêu hộp cung cấp thông tin về mức độ đại diện và độ ổn định. Việc đánh giá siêu hộp là cơ sở để kiểm soát nguồn thông tin được sử dụng trong học bán giám sát.

2.1.3. Cơ chế giám sát bán giám sát dựa trên siêu hộp

Thông tin nhân của các siêu hộp có nhãn được sử dụng để hỗ trợ gán nhãn cho các đối tượng chưa biết nhãn. Tuy nhiên, nếu quyết định chỉ dựa vào một siêu hộp, sai lệch của siêu hộp đó có thể ảnh hưởng trực tiếp đến kết quả. Đây là động lực để đưa cơ chế bỏ phiếu đa số vào FMN. [19, 21]

2.2. Học bán giám sát dựa trên bỏ phiếu đa số trong FMN

2.2.1. Giới thiệu về phương pháp bỏ phiếu đa số

Bỏ phiếu đa số tổng hợp quyết định của nhiều đối tượng thay vì phụ thuộc vào một nguồn duy nhất. Trong FMN, các đối tượng tham gia bỏ phiếu có thể là

các siêu hộp lân cận hoặc các mẫu có nhãn lân cận. Nhãn có số phiếu cao nhất được ưu tiên gán cho đối tượng chưa biết nhãn. [21]

2.2.2. Kỹ thuật bỏ phiếu đa số trong mạng nơron FMN

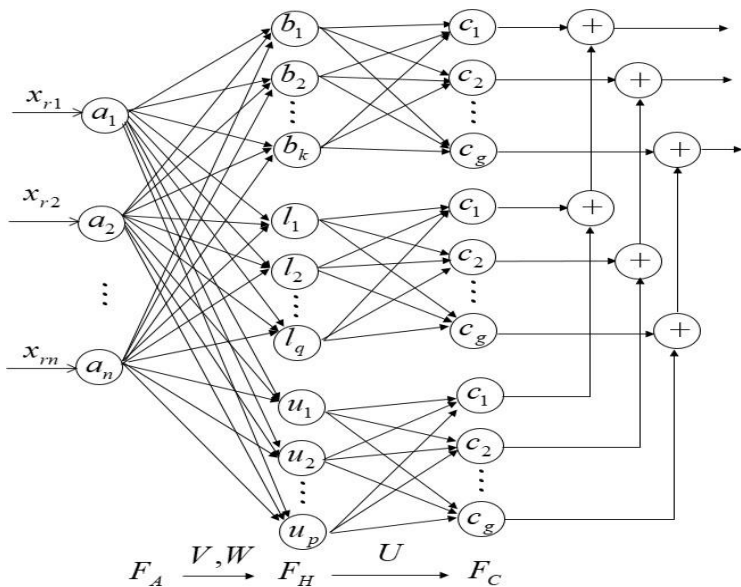
Kỹ thuật bỏ phiếu được tích hợp vào quá trình học bán giám sát sau khi cấu trúc siêu hộp đã được hình thành. Việc lựa chọn tập lân cận dựa trên độ thuộc hoặc quan hệ khoảng cách, sau đó tổng hợp nhãn của các đối tượng lân cận để đưa ra quyết định.

2.3. Mô hình FMN-Voting

2.3.1. Ý tưởng của FMN-Voting

FMN-Voting suy luận nhãn của mẫu chưa gán nhãn bằng cơ chế bỏ phiếu đa số từ các siêu hộp lân cận có độ thuộc cao. Thay vì sử dụng một siêu hộp duy nhất, mô hình khai thác đồng thời nhiều siêu hộp, nhờ đó tăng tính ổn định của quyết định trong các vùng dữ liệu có chồng lấn hoặc không chắc chắn. [21]

2.3.2. Kiến trúc mạng nơron FMN-Voting



Hình 2.3: Kiến trúc mạng nơron FMN-Voting

2.3.3. Thuật toán học trong FMN-Voting

Thuật toán học FMN-Voting gồm hai giai đoạn:

Giai đoạn 1 - Xây dựng tập siêu hộp hạt giống.

Từ tập mẫu có nhãn X_L , FMN xây dựng tập siêu hộp mang nhãn $B = B_{jj=1}^K$. Với mỗi $x_r \in X_L$, Lựa chọn B_j có cùng nhãn và: $\mu(x_r, B_j) = \max_{B_k \in B} \mu(x_r, B_k)$, đồng thời thỏa mãn: $\delta(x_r, B_j) \leq \delta^{max}$. Nếu thỏa mãn, B_j được mở rộng; ngược lại tạo siêu hộp mới $B_{new} = (x_r, x_r)$ và gán $\ell(B_{new}) = \ell(x_r)$. Kết quả thu được tập siêu hộp hạt giống có nhãn làm đầu vào cho giai đoạn 2.

Giai đoạn 2 - Học bán giám sát bằng bỏ phiếu đa số.

Với mỗi mẫu chưa gán nhãn $x_r \in X_U$, xác định tập k siêu hộp lân cận có độ thuộc $\mu(x_r, B_j)$ lớn nhất. Nhãn của x_r được suy luận theo nguyên tắc bỏ phiếu đa số:

$$\hat{\ell}(x_r) = \arg \max_c \sum_{B_j \in N_k(x_r)} I(\ell(B_j) = c), \text{ trong đó } N_k(x_r) \text{ là tập } k \text{ siêu hộp}$$

lân cận của x_r . Sau khi xác định $\hat{\ell}(x_r)$, mẫu được sử dụng để mở rộng siêu hộp phù hợp hoặc tạo siêu hộp mới nếu không thỏa mãn điều kiện $\delta(x_r, B_j) \leq \delta^{max}$.

Như vậy, Giai đoạn 1 tạo tập siêu hộp hạt giống từ X_L ; Giai đoạn 2 sử dụng các siêu hộp mang nhãn để bỏ phiếu, suy luận nhãn và học từ X_U , qua đó mở rộng dần thông tin giám sát trong FMN.

2.3.4. Độ phức tạp tính toán của FMN-Voting

Với $M = |X|$ là số mẫu, n là số thuộc tính và $K = |B|$ là số siêu hộp, chi phí xử lý mỗi mẫu bị chặn trên bởi $O(nK)$. Tổng chi phí của giai đoạn huấn luyện bán giám sát là $O(nMK)$. Do cả hai giai đoạn chính đều có cùng bậc độ phức tạp, tổng chi phí FMN-Voting vẫn ở bậc $O(nMK)$.

$$T = T_{_1} + T_{_2} = O(nMK) + O(MnK) \quad (2.22)$$

2.4. Mô hình FMN-SAL

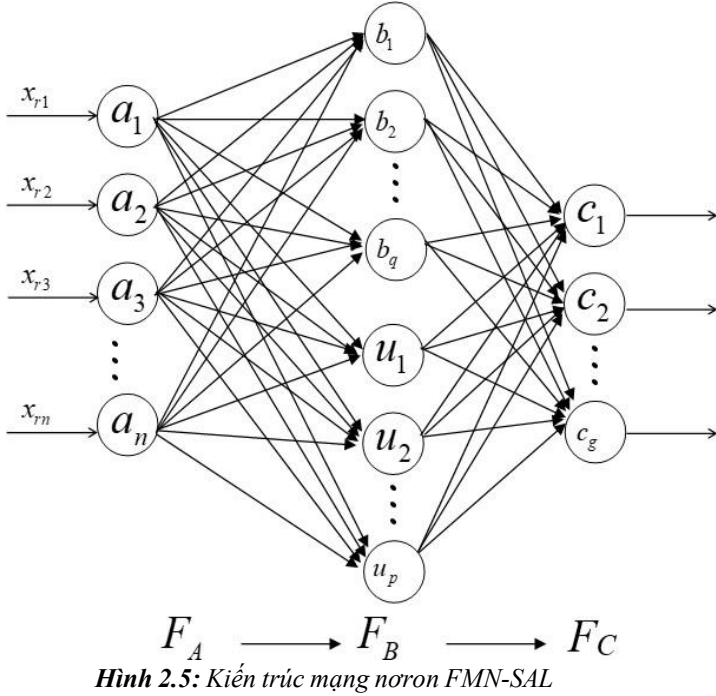
2.4.1. Ý tưởng của FMN-SAL

FMN-Voting giả định các siêu hộp lân cận có mức độ đại diện tương đương. Giả định này không luôn đúng khi dữ liệu trong siêu hộp phân bố không đều hoặc bị nhiễu. FMN-SAL khắc phục bằng cách khai thác trực tiếp các mẫu có

nhân lân cận để bỏ phiếu, giảm phụ thuộc vào chất lượng hình học của từng siêu hộp. [19, 21].

2.4.2. Kiến trúc mạng neuron FMN-SAL

FMN-SAL kế thừa cơ chế phân hoạch siêu hộp của FMN nhưng thay đổi nguồn thông tin dùng cho suy luận nhân. Tập các mẫu có nhãn lân cận được sử dụng để xác định nhân theo đa số, qua đó tăng khả năng khai thác thông tin cục bộ. [19]



Hình 2.5: Kiến trúc mạng neuron FMN-SAL

2.4.3. Thuật toán học trong FMN-SAL

Thuật toán thực hiện phân hoạch dữ liệu bằng FMN, duy trì các mẫu mang nhãn và khi cần suy luận cho đối tượng chưa gán nhãn sẽ xác định tập lân cận phù hợp, sau đó bỏ phiếu dựa trên nhãn của các mẫu lân cận. Cơ chế này nhằm hạn chế ảnh hưởng của các siêu hộp mất cân bằng hoặc không đại diện tốt cho phân bố dữ liệu. Thuật toán FMN-SAL thực hiện học bán giám sát trên tập $X = (x_r, \ell_{x_r})_{r=1}^M$ theo hai trường hợp chính. Với mỗi mẫu x_r , thuật toán chọn siêu hộp B_j có độ thuộc lớn nhất và thỏa mãn điều kiện mở rộng:

$\mu(x_r, B_j) = \max_{B_k \in B} \mu(x_r, B_k)$, $\delta(x_r, B_j) \leq \delta^{max}$. Nếu tồn tại B_j phù hợp, siêu hộp được điều chỉnh theo (1.6), (1.7); với mẫu chưa gán nhãn $\ell_{x_r} = 0$, mẫu được gán tạm nhãn của siêu hộp, $\ell_{x_r} \leftarrow \ell(B_j)$. Sau đó, các siêu hộp khác nhãn bị chồng lấn được kiểm tra và điều chỉnh. Nếu không tồn tại siêu hộp phù hợp, với $\ell(x_r) \neq 0$, thuật toán tạo siêu hộp mới B_{new} và gán $\ell(B_{new}) \leftarrow \ell_{x_r}$. Ngược lại, với mẫu chưa gán nhãn, thuật toán xác định tập lân cận $N(x_r)$ theo (2.23) và thực hiện bỏ phiếu trên các mẫu có nhãn lân cận. Khi $|N(x_r)| = k$, nhãn chiếm đa số được gán cho B_{new} theo (2.25).

Quá trình lặp đến khi $X = \emptyset$; sau đó cập nhật $X \leftarrow X'$ và cập nhật U dựa trên tập siêu hộp B theo (1.36). Bản chất của FMN-SAL là sử dụng các mẫu có nhãn lân cận thay cho các siêu hộp lân cận để bỏ phiếu suy luận nhãn, qua đó giảm ảnh hưởng của các siêu hộp có mức độ đại diện không đồng đều.

2.5. Thực nghiệm và đánh giá

2.5.1. Thiết lập thực nghiệm

Mục tiêu thực nghiệm. Thực nghiệm được thực hiện nhằm đánh giá hiệu quả của hai mô hình đề xuất FMN-Voting và FMN-SAL, tập trung vào khả năng phân cụm khi tỷ lệ dữ liệu có nhãn thấp, mức độ ổn định khi thay đổi tham số và so sánh với các phương pháp đối sánh trên các tập dữ liệu có cấu trúc khác nhau.

Phương pháp thực nghiệm. Các mô hình được đánh giá trên các tập dữ liệu chuẩn Aggregation, Jain, R15, Flame và các tập dữ liệu phân loại Ecoli, Soybean, Zoo, Yeast, đồng thời thực nghiệm trên dữ liệu ILPD. Chất lượng mô hình được đánh giá bằng các chỉ số RI, ARI, NMI đối với phân cụm và Accuracy, Precision, Recall, Specificity, F1-Score đối với bài toán có nhãn. Kết quả của FMN-Voting và FMN-SAL được so sánh với các mô hình FMN và các phương pháp đối sánh được sử dụng trong luận án.

Thiết lập tham số. Các tham số chính gồm hệ số mờ γ , kích thước siêu hộp cực đại δ^{max} , ngưỡng tin cậy β và số lượng đối tượng lân cận k phục vụ bỏ phiếu.

Thực nghiệm thay đổi δ^{max} và tỷ lệ mẫu có nhãn để khảo sát ảnh hưởng đến số lượng siêu hộp và chất lượng phân cụm; các tham số còn lại được thiết lập theo từng kịch bản thực nghiệm như quy định trong luận án.

2.5.2. Kết quả thực nghiệm và đánh giá

Trên các tập dữ liệu chuẩn, FMN-SAL cho kết quả ARI cao và ổn định. Chẳng hạn, trên Aggregation, Jain, R15 và Flame, ARI của FMN-SAL lần lượt

đạt khoảng 0.9911; 0.9879; 0.9980 và 0.9798. Các kết quả cho thấy cơ chế khai thác mẫu có nhãn lân cận phù hợp với các tập dữ liệu có cấu trúc cụm phức tạp.

Bảng tóm tắt kết quả ARI tiêu biểu (trích Bảng 2.4 của luận án)

<i>Mô hình</i>	<i>Aggregation</i>	<i>Jain</i>	<i>R15</i>	<i>Flame</i>
GCAE	0.992	1	0.9928	0.9501
DBSCAN	0.9664	0.9382	0.9375	0.9388
FMN-Voting	-	-	0.97804	0.96218
FMN-SAL	0.991139	0.987856	0.998	0.979817

Trên dữ liệu ILPD, FMN-SAL đạt Accuracy 0.8806 ± 0.0043 , Specificity 0.8001 ± 0.0082 , Precision 0.8786 ± 0.0044 và F1-Score 0.8788 ± 0.0043 . FMN-Voting đạt Accuracy 0.8045 ± 0.0366 , cao hơn SCFMN; các phương pháp K-means++, EM và DBSCAN có Accuracy thấp hơn các mô hình FMN đề xuất.

Bảng 2.12: So sánh kết quả của FMN-SAL, FMN-Voting và các phương pháp đối sánh.

Mô hình	Accuracy	Recall	Specificity	FP Rate	Precision	F1-Score
K-means++	0.6252 ± 0.0222	0.6252 ± 0.0222	0.4463 ± 0.0932	0.5537 ± 0.0932	0.6148 ± 0.0437	0.6154 ± 0.0240
EM	0.5647 ± 0.0351	0.5647 ± 0.0351	0.7353 ± 0.1299	0.2647 ± 0.1299	0.7589 ± 0.0544	0.5648 ± 0.0249
DBSCAN	0.6416 ± 0.0022	0.6416 ± 0.0022	0.4411 ± 0.0048	0.5589 ± 0.0048	0.6229 ± 0.0021	0.6306 ± 0.0019
SCFMN	0.7780 ± 0.0184	0.7780 ± 0.0184	0.6442 ± 0.0330	0.3558 ± 0.0330	0.7718 ± 0.0185	0.7736 ± 0.0186
FMN-SAL	0.8806 ± 0.0043	0.8806 ± 0.0043	0.8001 ± 0.0082	0.1999 ± 0.0082	0.8786 ± 0.0044	0.8788 ± 0.0043
FMN-Voting	0.8045 ± 0.0366	0.8045 ± 0.0366	0.7029 ± 0.0490	0.2971 ± 0.0490	0.8021 ± 0.0365	0.8030 ± 0.0364

Kết luận chương 2

Chương 2 đã phát triển hai mô hình FMN-Voting và FMN-SAL theo hướng tăng độ tin cậy của suy luận nhãn bằng bỏ phiếu. FMN-Voting tổng hợp nhãn từ các siêu hộp lân cận; FMN-SAL khai thác các mẫu có nhãn lân cận để giảm ảnh hưởng của siêu hộp kém đại diện. Kết quả thực nghiệm cho thấy hai mô hình cải thiện chất lượng phân cụm và phân loại, trong đó FMN-SAL thể hiện độ ổn định và khả năng khái quát tốt trên nhiều dạng dữ liệu.

Một phần kết quả Chương 2 được công bố trong [CT2], [CT3], [CT4] và [CT8].

CHƯƠNG 3

HỌC BÁN GIÁM SÁT TRONG MẠNG NƠON FMN VỚI CHIẾN LƯỢC CHỌN HẠT GIỐNG CHỦ ĐỘNG

3.1. Giới thiệu về mô hình FA-Seed

FMN-Voting và FMN-SAL cải thiện suy luận nhãn nhưng hiệu quả của quá trình học vẫn phụ thuộc đáng kể vào chất lượng hạt giống ban đầu. Trong không gian dữ liệu phức tạp, các hạt giống không có vai trò tương đương; các mẫu ở vùng biên, vùng nhiễu hoặc vùng chồng lấn có thể làm lan truyền lỗi nhãn.

FA-Seed được xây dựng theo hướng lựa chọn hạt giống chủ động và linh hoạt, ưu tiên các đối tượng có tính đại diện và độ tin cậy cao. Mục tiêu là sử dụng hiệu quả ngân sách nhãn hạn chế và tạo tập hạt giống có khả năng dẫn dắt tốt quá trình phân cụm bán giám sát. [102]

3.2. Mô hình FA-Seed: Lựa chọn hạt giống chủ động dựa trên FMN

3.2.1. Mô tả bài toán

Cho tập dữ liệu X gồm các mẫu trong không gian n chiều, trong đó chỉ một phần nhỏ có thể được truy vấn/gán nhãn. Bài toán đặt ra là lựa chọn các mẫu hoặc siêu hộp phù hợp để hình thành tập hạt giống, sau đó sử dụng tập hạt giống này cho quá trình học bán giám sát.

3.2.2. Ý tưởng của FA-Seed

FA-Seed sử dụng cấu trúc siêu hộp FMN để tổ chức dữ liệu và đánh giá các ứng viên hạt giống. Thay vì lựa chọn ngẫu nhiên, mô hình xem xét mức độ đại diện và độ tin cậy của từng hạt giống. Cách tiếp cận thích nghi giúp ưu tiên các vùng dữ liệu có khả năng cung cấp thông tin tốt, đồng thời hạn chế các đối tượng không ổn định.

Quá trình lựa chọn được gắn với cơ chế học chủ động: mô hình xác định các ứng viên có giá trị thông tin, truy vấn nhãn trong giới hạn ngân sách và cập nhật tập hạt giống. Sau đó, thông tin từ các hạt giống được sử dụng để dẫn dắt quá trình lan truyền nhãn và hình thành cụm. [102]

Độ lệch tâm của siêu hộp: $e(B_j) = \|g(B_j) - d(B_j)\|_2$ (3.8)

Trong đó $g(B_j)$ là tâm hình học, còn $d(B_j)$ là tâm dữ liệu của siêu hộp. Giá trị $e(B_j)$ càng nhỏ thì phân bố dữ liệu trong siêu hộp càng cân đối.

Tập hạt giống mở rộng từ các siêu hộp được lựa chọn:

$$S = \bigcup_{j \in J_Q} \{x_r \in X \mid \mu(x_r, B_j) \geq \tau\}, \quad \tau \in [0, 1] \quad (3.9)$$

Tập lõi của siêu hộp B_j : $T_j = \{x \mid \mu_j(x_r, B_j) \geq \tau\}$ (3.10)

Với $\tau = 1$, tập T_j gồm các mẫu có độ thuộc đầy đủ vào B_j .

Tập chỉ số các siêu hộp thuần khiết: $J_Q = \{j | B_j \in B, e(B_j) \leq \varepsilon\}$ (3.11)

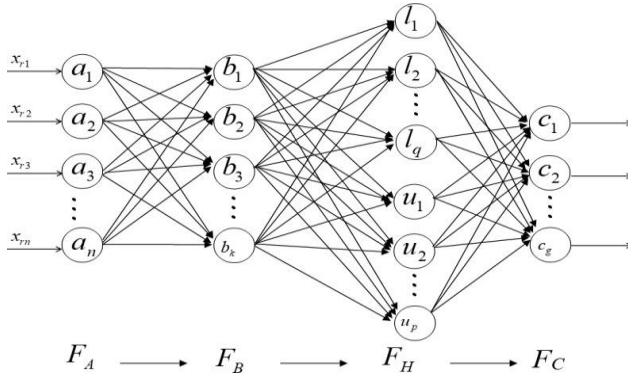
Đây là điều kiện lựa chọn siêu hộp ứng viên quan trọng nhất: chỉ những siêu hộp có độ lệch tâm không vượt quá ngưỡng ε mới được xem là đủ thuần khiết để tham gia sinh hạt giống.

Tập ứng viên hạt giống: $Q = \{d(B_j) | j \in J_Q\}$ (3.12)

Trong đó mỗi phần tử của Q là tâm dữ liệu $d(B_j)$ của một siêu hộp thuần khiết và được sử dụng làm đối tượng truy vấn nhãn từ chuyên gia.

3.2.3. Kiến trúc mạng nơron FA-Seed

Kiến trúc FA-Seed kế thừa lớp biểu diễn siêu hộp của FMN và bổ sung cơ chế đánh giá, lựa chọn hạt giống để phục vụ học bán giám sát. Sự tách biệt giữa xây dựng cấu trúc siêu hộp và lựa chọn nguồn giám sát giúp mô hình kiểm soát tốt hơn chất lượng thông tin nhãn.



Hình 3.9: Kiến trúc mạng nơron FA-Seed

3.2.4. Thuật toán học của mô hình FA-Seed

Thuật toán FA-Seed gồm các bước chính: phân hoạch dữ liệu bằng FMN; đánh giá các siêu hộp/ứng viên; lựa chọn chủ động các hạt giống theo tiêu chí đại diện và độ tin cậy; truy vấn hoặc tiếp nhận nhãn; sau đó thực hiện học bán giám sát và cập nhật cấu trúc mô hình. Cơ chế thích nghi cho phép mô hình điều chỉnh tập hạt giống theo trạng thái của dữ liệu thay vì cố định ngay từ đầu.

Thuật toán 3.2. Huấn luyện bán giám sát có dẫn dắt

Input: $X; B; \beta; \delta^{max}$

Output: $B; U; X$

```

1:   $B_{lab} = \emptyset; B_{unl} = \emptyset;$ 
2:  for mỗi  $x_r \in X$  do
3:      Với  $B_j \in B_{lab} \cup B_{unl}$ , chỉ số  $j^*$  được xác định theo (3.19);
4:      if  $x_r \subseteq B_{j^*}$  then
5:          Mở rộng siêu hộp  $B_{j^*}$  theo Công thức (3.21);
6:      if ( $B_{j^*}$  có chồng lấn với  $B_k$ ) với  $\ell(B_{j^*}) \neq \ell(B_k)$  then
7:          Điều chỉnh  $B_{j^*}$  và  $B_k$  theo các công thức (1.15) đến (1.26);
8:      end if
9:      else
10:         Tạo  $B_{new}$ , khởi tạo các điểm  $V_{new}, W_{new}$  bởi  $x_r$ ;
11:         if  $\ell_{x_r} \neq 0$  then
12:              $\ell(B_{new}) = \ell_{x_r}; B_{lab} = B_{lab} \cup B_{new};$ 
13:         else
14:              $B_{unl} = B_{unl} \cup B_{new};$ 
15:         end if
16:     end if
17: end for
18: repeat
19:     for mỗi  $B_j \in B_{unl}$  do
20:         Tính trọng tâm  $g(B_j)$  theo Công thức (1.40);
21:         if ( $\exists B_k^*$  với  $B_k \in B_{lab}$  thỏa mãn (3.22)) then
22:              $\ell(B_j) = \ell(B_k); B_{lab} = B_{lab} \cup B_j; B_{unl} = B_{unl} \setminus B_j;$ 
23:         end if
24:     end for
25:     Giảm  $\beta$ ;
26: until ( $B_{unl} = \emptyset$ ) or ( $\beta \leq \text{ngưỡng tin cậy}$ );
27: if  $B_{unl} \neq \emptyset$  then
28:     for mỗi  $B_j \in B_{unl}$  do

```

- 29: Tính trọng tâm $g(B_j)$ theo Công thức (1.40);
- 30: Đặt $k^* \leftarrow \arg \max_k, \mu(g(B_j), B_k)$ với $B_k \in B_{lab}$;
 $\ell(B_j) = \ell(B_{k^*})$;
- 31: **end for**
- 32: **end if**
- 33: $B = B_{lab} \cup B_{unl}$;
- 34: **return** B .

Độ phức tạp tính toán của FA-Seed được xác định là:

$$T = T_1 + T_2 = O(nMK) + O(nMK) = O(nMK). \quad (3.27)$$

Trong đó $M = |X|$ là số mẫu, n là số thuộc tính và $K = |B|$ là số siêu hợp. Trong thực tế, K thường nhỏ hơn đáng kể so với M , nhờ đó FA-Seed có khả năng mở rộng khi kích thước dữ liệu tăng.

3.3. Thực nghiệm và đánh giá

3.3.1. Thiết lập thực nghiệm

FA-Seed được đánh giá trên các tập dữ liệu chuẩn và dữ liệu bệnh gan, đồng thời so sánh với MSCFMN và các phương pháp phân cụm khác. Các thí nghiệm xem xét ảnh hưởng của δ^{max} , khả năng nhận diện từng lớp, chất lượng phân hoạch và thời gian chạy. [102]

3.3.2. Kết quả thực nghiệm và đánh giá

Kết quả cho thấy chiến lược lựa chọn hạt giống chủ động giúp FA-Seed cải thiện chất lượng phân cụm so với các phương pháp đối sánh trong nhiều kịch bản. Mô hình không chỉ hướng tới độ chính xác mà còn chú trọng độ ổn định của nguồn thông tin giám sát, đặc biệt khi dữ liệu có nhiễu, chồng lấn hoặc phân bố không đồng đều.

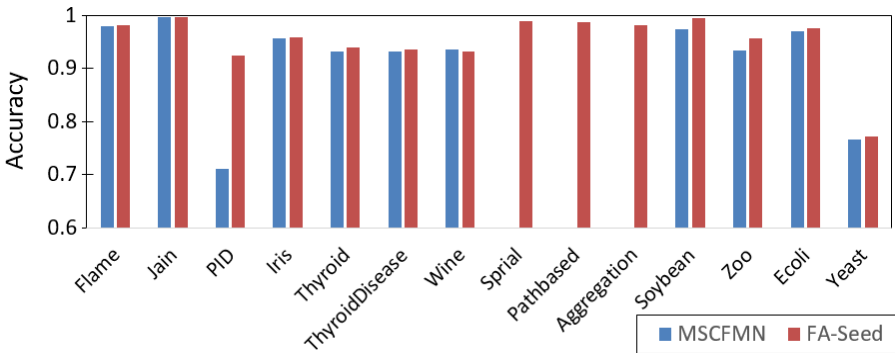
Trên dữ liệu bệnh gan, FA-Seed đạt các chỉ số phân hoạch cao hơn rõ rệt so với MSCFMN và các phương pháp K-means++, EM, DBSCAN. Bảng sau giữ lại kết quả tiêu biểu từ Bảng 3.7 của luận án.

Mô hình	RI	ARI	NMI
K-means++	0.5314 ± 0.0111	0.0189 ± 0.0336	0.0107 ± 0.0110
EM	0.5097 ± 0.0146	-0.0130 ± 0.0144	0.0926 ± 0.0343
DBSCAN	0.5393 ± 0.0013	0.0347 ± 0.0030	0.0066 ± 0.0011

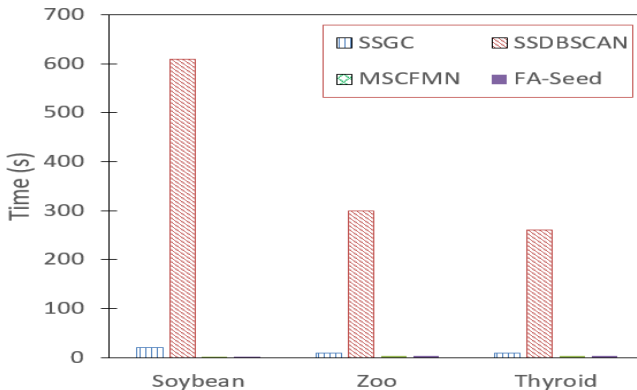
MSCFMN	0.6783 ± 0.0402	0.3279 ± 0.0812	0.1993 ± 0.0665
FA-Seed	0.8215 ± 0.0263	0.6285 ± 0.0549	0.4779 ± 0.0577

FA-Seed đạt RI 0.8215 ± 0.0263 , ARI 0.6285 ± 0.0549 và NMI 0.4779 ± 0.0577 , cao hơn MSCFMN và các phương pháp trên Weka. Kết quả cho thấy tập hạt giống được lựa chọn phù hợp đã hỗ trợ quá trình phân cụm tạo ra phân hoạch gần với nhãn thực tế hơn.

Phân tích thời gian chạy cho thấy FA-Seed có lợi thế so với các phương pháp cơ sở trong các thí nghiệm được báo cáo. Với độ phức tạp $O(nMK)$, mô hình tận dụng thực tế số siêu hộp K thường nhỏ hơn số mẫu M . Trong khi đó, các phương pháp dựa trên đồ thị có thể phát sinh chi phí lớn khi xây dựng quan hệ láng giềng trên dữ liệu quy mô lớn. [103]



Hình 3.14: Biểu đồ so sánh trực quan chỉ số Accuracy giữa FA-Seed và MSCFMN.



Hình 3.15: So sánh thời gian chạy (s) giữa FA-Seed và các thuật toán cơ sở trên các tập dữ liệu thực nghiệm

Các kết quả thực nghiệm củng cố nhận định rằng chất lượng hạt giống là thành phần quan trọng của học bán giám sát. Khi ưu tiên các hạt giống có mức độ đại diện và độ tin cậy tốt, nguy cơ lan truyền lỗi nhãn được giảm và chất lượng phân cụm được cải thiện.

Kết luận chương 3

Chương 3 đã trình bày FA-Seed với trọng tâm là lựa chọn hạt giống chủ động. Mô hình không giả định mọi hạt giống có vai trò tương đương mà xem xét đồng thời mức độ đại diện và độ tin cậy. Việc lựa chọn thích nghi giúp giảm nguy cơ lan truyền lỗi nhãn trong vùng chồng lấn, nhiễu hoặc phân bố không đồng đều.

Phân tích độ phức tạp và thực nghiệm cho thấy FA-Seed bảo đảm tính khả thi và có tiềm năng cải thiện chất lượng phân cụm. Một phần kết quả Chương 3 đã được công bố trong [CT5], [CT6] và [CT7].

KẾT LUẬN

A. Những kết quả chính của luận án

Luận án tập trung nghiên cứu phân cụm bán giám sát trong mạng neuron FMN trong bối cảnh dữ liệu ngày càng lớn và phức tạp, trong khi thông tin giám sát khan hiếm. Trên cơ sở tổng quan có hệ thống, luận án xác định hạn chế của các mô hình FMN bán giám sát khi suy luận và lan truyền nhãn phụ thuộc vào các thực thể đơn lẻ hoặc khi chất lượng hạt giống không được kiểm soát.

Kết quả thứ nhất là mô hình FMN-Voting, đưa cơ chế bỏ phiếu đa số dựa trên các siêu hộp lân cận vào quá trình suy luận nhãn. Cơ chế này làm giảm sự phụ thuộc vào một siêu hộp đơn lẻ và tăng độ ổn định của quyết định.

Kết quả thứ hai là FMN-SAL, khai thác trực tiếp các mẫu mang nhãn lân cận để bỏ phiếu. Cách tiếp cận này giảm ảnh hưởng của các siêu hộp có chất lượng hoặc mức độ đại diện không đồng đều và cho kết quả tốt trên nhiều tập dữ liệu.

Kết quả thứ ba là FA-Seed, phát triển chiến lược lựa chọn hạt giống chủ động dựa trên cấu trúc FMN, có xét đến mức độ đại diện và độ tin cậy. FA-Seed giúp hình thành tập hạt giống phù hợp hơn và hạn chế nguy cơ lan truyền lỗi nhãn.

B. Những đóng góp mới của luận án

Các đóng góp của luận án mở rộng FMN bán giám sát theo hướng ổn định và tin cậy hơn: từ tổng hợp thông tin của nhiều siêu hộp, chuyển sang khai thác trực tiếp mẫu có nhãn lân cận, và cuối cùng chủ động kiểm soát chất lượng nguồn giám sát thông qua lựa chọn hạt giống.

C. Hạn chế và hướng nghiên cứu tiếp theo

Mặc dù các mô hình đề xuất đạt kết quả tích cực, việc xử lý dữ liệu có cấu trúc đặc trưng phức tạp và dữ liệu quy mô rất lớn vẫn cần tiếp tục nghiên cứu. Hiệu quả của mô hình cũng còn phụ thuộc vào biểu diễn đặc trưng, cấu trúc siêu hộp và cơ chế suy luận nhãn.

Các hướng nghiên cứu tiếp theo của luận án gồm: (i) kết hợp FMN bán giám sát với các mô hình học sâu hoặc biểu diễn đặc trưng để xử lý dữ liệu có cấu trúc phức tạp; (ii) phát triển cơ chế suy luận nhãn thích nghi hơn dựa trên độ tin cậy và động học của siêu hộp; (iii) nghiên cứu chiến lược song song và phân tán để nâng cao khả năng mở rộng trên tập dữ liệu lớn; và (iv) tiếp tục khai thác khả năng kết xuất luật mờ nhằm tăng tính diễn giải và khả năng ứng dụng trong các hệ thống hỗ trợ ra quyết định.

Nhìn chung, luận án đóng góp các kết quả mới về phương pháp và ứng dụng cho bài toán phân cụm bán giám sát trong FMN, góp phần phát triển hướng nghiên cứu này theo hướng ổn định hơn, tin cậy hơn và phù hợp hơn với dữ liệu thực tế trong điều kiện nhãn hạn chế.

DANH MỤC CÔNG TRÌNH CÔNG BỐ LIÊN QUAN ĐẾN LUẬN ÁN

CT1 Nguyễn Thanh Sơn, Vũ Thị Dương, Nguyễn Anh Thái, Hoàng Quang Huy, Vũ Đình Minh. “Tổng quan về mạng nơron min-max mờ và ứng dụng”, Một số vấn đề chọn lọc của Công nghệ thông tin và Truyền thông, Hội thảo Quốc gia lần thứ XXV, Hà Nội, 8-9/12/2022, NXB Khoa học và Kỹ thuật, 2022, tr. 238-244.

CT2 Dinh-Minh Vu, Thanh-Son Nguyen, Trung-Nghia Phung, Trinh-Hoang Vu. “Choosing Data Points to Label for Semi-supervised Learning Based on Neighborhood”, Advances in Information and Communication Technology: Proceedings of ICTA 2023, Lecture Notes in Networks and Systems, vol. 847, Springer, 2023, tr. 134-141.

CT3 Dinh-Minh Vu, Thanh-Son Nguyen, Van Tinh Nguyen, Duc-Luu Nguyen. “FMN-Voting: A Semi-supervised Clustering Technique Based on Voting Methods Using FMM Neural Networks”, Advances in Information and Communication Technology: Proceedings of the 3rd International Conference ICTA 2024, Lecture Notes in Networks and Systems, vol. 1205, Springer Cham, 2025, tr.137-150. DOI: 10.1007/978-3-031-80943-915.

CT4 Thanh-Son Nguyen, Dinh-Minh Vu, Duc-Luu Nguyen. “FMN-SSL: A Semi-supervised Approach for Lung Tumor Segmentation”, technical program ICTA 2025; proceedings Advances in Information and Communication Technology: Proceedings of the 4th International Conference, ICTA 2025, Volume 3, Lecture Notes in Networks and Systems, vol. 1833, Springer Cham, phát hành online tháng 3/2026.

CT5 Dinh-Minh Vu, Thanh-Son Nguyen, Trinh-Hoang Vu, Thi-Duong Vu, Duc-Luu Nguyen, Quang-Huy Hoang, Ba-Dung Le. “FMN-OE: Evaluation and Elimination of Hyperbox Overlap”, Intelligent Systems and Data Science: Third International Conference, ISDS 2025, Proceedings, Part II, Communications in Computer and Information Science, vol. 2714, Springer, Singapore, 2026, tr. 328-342. DOI: 10.1007/978-981-95-3358-9_28.

CT6 Dinh-Minh Vu, Thanh-Son Nguyen. “FA-Seed: Flexible and Active Learning-Based Seed Selection”. Information, 2025, 16, 884. DOI: 10.3390/info16100884. (ESCI, Scopus).

CT7 Nguyễn Thanh Sơn, Vũ Đình Minh, Lê Bá Dũng, Nguyễn Văn Thanh. “Phân cụm bán giám sát với phương pháp chọn hạt giống dựa trên trọng số siêu hộp trong mạng nơron min-max mờ”, Một số vấn đề chọn lọc của Công nghệ

thông tin và Truyền thông, Hội thảo Quốc gia lần thứ XXIV, Thái Nguyên, 13-14/12/2021, NXB Khoa học và Kỹ thuật, 2021, tr. 433-438.

CT8 Nguyễn Thanh Sơn, Vũ Đình Minh. “Integrating Active Learning And Controlled Label Propagation In Fuzzy Min-Max Neural Networks”, Journal of Computer Science and Cybernetics. Công trình [CT8] hiện đã hoàn thành quá trình phản biện vòng 2 tại Tạp chí Tin học và Điều khiển học.