

**BỘ GIÁO DỤC
VÀ ĐÀO TẠO**

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

HỌC VIỆN KHOA HỌC VÀ CÔNG NGHỆ



HUỲNH THỊ NGỌC CẨM

**TIẾP CẬN HỌC MÁY ĐỂ TUYỂN CHỌN
VÀ TRỰC QUAN HÓA ĐỂ HUẤN LUYỆN
HỌC SINH ĐỘI TUYỂN**

TÓM TẮT LUẬN ÁN TIẾN SĨ MÁY TÍNH

Ngành: Khoa học máy tính

Mã số: 9 48 01 01

Hà Nội - 2025

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:

PGS.TS. Trần Vĩnh Phước, Trường Đại học Tài nguyên và Môi trường TP.HCM

Phản biện 1:.....

.....

Phản biện 2:.....

.....

Phản biện 3:.....

.....

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 giờ , ngày tháng năm

Có thể tìm hiểu luận án tại:

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

I. MỞ ĐẦU

1. Lý do chọn đề tài

Mỗi năm, Bộ Giáo dục Đào tạo Việt Nam đều tổ chức thi học sinh giỏi cấp quốc gia dành cho học sinh các trường trung học phổ thông tham gia theo từng môn học như Tin học, Toán học, Vật lý, Hóa học, Sinh học, v.v... Để tổ chức cho học sinh tham gia kỳ thi học sinh giỏi, các trường trung học phổ thông thành lập đội tuyển cho mỗi môn thi và giao cho một giáo viên phụ trách với chức danh huấn luyện viên. Mỗi huấn luyện viên có nhiệm vụ tuyển chọn đội viên và huấn luyện chuyên sâu theo yêu cầu của kỳ thi trước khi tham gia cuộc thi. Trước yêu cầu về thành tích của trường, mỗi huấn luyện viên được đặt trước các thách thức:

- Làm thế nào hiểu được các yếu tố năng lực cần có của một thí sinh tham dự kỳ thi học sinh giỏi cấp quốc gia của một môn học nào đó?
- Làm thế nào để tuyển chọn đúng học sinh giỏi có năng lực đáp ứng đúng yêu cầu của kỳ thi học sinh giỏi cấp quốc gia?
- Áp dụng phương pháp nào để huấn luyện nâng cao năng lực của đội viên đáp ứng đúng yêu cầu của đề thi học sinh giỏi cấp quốc gia?

Về phương pháp huấn luyện, Bộ Giáo dục Đào tạo đang đẩy mạnh dự án STEM trong các chương trình phổ thông nhằm hướng tới giảng dạy kết hợp các môn khoa học (science), công nghệ (technology), kỹ thuật (engineering), và toán (mathematic) với phương pháp đặt người học vào trung tâm quá trình dạy và học. Mục tiêu của STEM không phù hợp với việc huấn luyện đội tuyển theo từng môn học. Học khám phá có hướng dẫn (guided-discovery learning) là tiếp cận phù hợp để huấn luyện đội tuyển vì đội viên được phát triển tính độc lập trong tư duy và hoạt động học tập để tham gia kỳ thi học sinh giỏi cấp quốc gia.

2. Mục tiêu của luận án

Mục tiêu chung của luận án: Luận án xây dựng quy trình tuyển chọn đội viên và huấn luyện đội tuyển học sinh giỏi môn Tin học của một trường trung học phổ thông để tham dự kỳ thi học sinh giỏi cấp quốc gia với kỳ vọng thắng được nhiều giải, dựa trên tiếp cận học máy và trực quan hóa dữ liệu trong khoa học máy tính.

Mục tiêu cụ thể:

- Phân tích các yếu tố năng lực của những thí sinh đã thắng giải (winners) và những thí sinh đã không thắng giải (nonwinners) để mô hình hóa toán học mỗi người như một vector toán học trong không gian nhiều chiều, mỗi chiều biểu diễn một biến đặc trưng năng lực.
- Xây dựng thuật toán và quy trình tuyển chọn học sinh giỏi vào đội tuyển học sinh giỏi môn Tin học của một trường trung học phổ thông để tham dự kỳ thi học sinh giỏi cấp quốc gia môn Tin học.
- Xây dựng mô hình học khám phá có hướng dẫn bằng kỹ thuật trực quan có đánh giá hiệu năng của từng bài giảng đối với đội tuyển, giáo viên.

3. Đối tượng nghiên cứu của luận án

- Năng lực thí sinh các trường trung học phổ thông đã từng tham gia giải học sinh giỏi cấp quốc gia môn Tin học.
- Dữ liệu giáo dục (Educational Data) về năng lực học sinh.
- Các tiếp cận học máy (machine learning) để khai phá dữ liệu giáo dục (EDM: Educational Data Mining)
- Lý thuyết học tập và mô hình học khám phá (DL: Discovery Learning)
- Nguyên lý tiếp nhận và kiến tạo kiến thức của con người (học sinh).

4. Phạm vi nghiên cứu của luận án

- Dữ liệu nghiên cứu liên quan thí sinh thắng giải và thí sinh không thắng giải cấp quốc gia môn Tin học.
- Đặc trưng năng lực học sinh gồm những đặc trưng học tập (learning features) và những đặc trưng phi học tập (non-learning features).
- Mô hình học khám phá có hướng dẫn - áp dụng thử nghiệm cho đội tuyển học sinh giỏi môn Tin học của một trường trung học phổ thông.

5. Phương pháp nghiên cứu

- *Phương pháp phân tích – tổng hợp.* Phương pháp phân tích được áp dụng để phân tích năng lực của mỗi học sinh như là một vector đặc trưng năng lực nhiều biến. Phương pháp phân tích - tổng hợp được áp dụng để cấu trúc bài hướng dẫn cho học sinh theo mô hình học khám phá có hướng dẫn trực quan.
- *Phương pháp mô hình hóa toán học.* Phương pháp mô hình hóa toán học được áp dụng để mô hình hóa năng lực của một học sinh như một vector đặc trưng năng lực trong không gian năng lực nhiều chiều.
- *Phương pháp học máy.* Luận án áp dụng phương pháp học máy có giám sát và đề xuất thuật toán Winner-domain để phân loại ứng viên, tìm ra những ứng viên có tiềm năng thắng giải quốc gia.
- *Phương pháp cây quyết định.* Luận án áp dụng phương pháp cây quyết định với thuật toán tương đồng cosin (cosine similarity) để quyết định chọn những ứng viên được kỳ vọng thắng giải cao vào đội tuyển.
- *Phương pháp toán học.* Phương pháp toán được áp dụng để xây dựng tập dữ liệu dataset.

- *Phương pháp lập trình.* Phần mềm Python được sử dụng để thiết lập các chương trình áp dụng cho công trình.
- *Phương pháp khảo sát.* Phương pháp khảo sát được áp dụng để xây dựng bộ câu hỏi (Questionnaire) và web Google Forms được sử dụng để thu thập dữ liệu.
- *Phương pháp thực nghiệm.* Phương pháp thực nghiệm được áp dụng cho thuật toán Winner-domain và thuật toán Winner-cosin với cây quyết định để chọn đội viên.

6. Các đóng góp của luận án

6.1. Kết quả nghiên cứu nội dung 1

- Tiếp cận biên soạn câu hỏi khảo sát định tính, đơn giản.
- Tiếp cận thành lập các biến đặc trưng năng lực và định lượng các hồi đáp khảo sát định tính.
- Dataset những thí sinh đã tham gia kỳ thi học sinh giỏi cấp quốc gia môn Tin học.

6.2. Kết quả nghiên cứu nội dung 2

Luận án đã đề xuất thuật toán Winner-domain xác định vùng năng lực của các winners trong không gian năng lực nhiều chiều, thuật toán Winner-cosin xác định tương đồng hướng năng lực của từng ứng viên được chọn bởi Winner-domain với tất cả winners trong winner dataset và cây quyết định xác định ứng viên được chọn thành đội viên.

6.3. Kết quả nghiên cứu nội dung 3

Luận án xây dựng mô hình học khám phá có hướng dẫn trực quan và đề xuất tiếp cận đánh giá hiệu năng của mô hình học khám phá có hướng dẫn bằng kỹ thuật trực quan đối với từng bài giảng của thầy hướng dẫn.

7. Cấu trúc luận án

Luận án được thiết kế gồm phần mở đầu, năm chương nội dung, và phần kết luận. Cuối luận án là danh mục các công trình của tác giả

được công bố quốc tế, danh mục các tài liệu tham khảo, và phụ lục thực nghiệm. Cụ thể, nội dung của các phần được tóm tắt như sau:

Phần mở đầu: Luận án giới thiệu bối cảnh phát sinh đề tài với 3 thách thức để nghiên cứu.

Chương 1 trình bày tổng quan các công trình về khai phá dữ liệu giáo dục được liên quan trong nghiên cứu.

Chương 2 trình bày cơ sở lý thuyết học khám phá.

Chương 3 thiết kế dataset năng lực thí sinh theo quy trình: phân tích đặc trưng năng lực \rightarrow mô hình hóa toán học năng lực \rightarrow thiết kế bộ câu hỏi khảo sát \rightarrow thu thập hồi đáp định tính \rightarrow lượng hóa các hồi đáp \rightarrow định lượng các tập biến năng lực của không gian năng lực và định lượng các đặc trưng năng lực của mỗi người được khảo sát \rightarrow thiết lập dataset. Luận án đã biên soạn câu hỏi định tính và đề xuất tiếp cận ma trận hồi đáp, khối 3D hồi đáp để lượng hóa các biến đặc trưng năng lực từ các hồi đáp định tính của người được khảo sát.

Chương 4 trình bày tiếp cận Winner-domain để thành lập miền tiềm năng và cây quyết định với tiếp cận Winner-cosine (tương đồng cosin) chọn ứng viên có năng lực trong miền tiềm năng (Winner-domain tốt).

Chương 5 trình bày tiếp cận học khám phá có hướng dẫn trực quan để huấn luyện đội tuyển.

Phần kết luận: Luận án tóm tắt các đóng góp chính của tác giả trong luận án và những nội dung sẽ được phát triển tiếp.

II. NỘI DUNG LUẬN ÁN

CHƯƠNG 1. TỔNG QUAN VỀ CÁC CÔNG TRÌNH KHAI PHÁ DỮ LIỆU GIÁO DỤC ĐƯỢC LIÊN QUAN

Những công trình khoa học liên quan chủ đề luận án là những nghiên cứu về năng lực và dữ liệu năng lực học sinh, những thuật toán

phân hoạch đối tượng thành 2 nhóm theo đặc trưng. Các công trình nghiên cứu được liên quan theo các chủ đề:

- Dữ liệu giáo dục.
- Khai phá dữ liệu giáo dục như support vector machine (SVM), cây quyết định, phân cụm K-mean, xử lý dữ liệu.

CHƯƠNG 2. CƠ SỞ LÝ THUYẾT HỌC KHÁM PHÁ

Luận án trình bày các cơ sở lý thuyết về các phương pháp giáo dục, các khái niệm học và học khám phá, các mô hình và phương pháp học khám phá, các tính chất của học khám phá và học khám phá có hướng dẫn. Cuối cùng là các lý thuyết về thu thập thông tin/tri thức của con người để đề xuất mô hình học khám phá có hướng dẫn trực quan tại chương 5.

CHƯƠNG 3. DATASET NĂNG LỰC THÍ SINH THAM GIA CÁC KỶ THI HỌC SINH GIỎI

3.1. Giới thiệu

3.2. Các yếu tố năng lực thí sinh

3.3. Câu hỏi khảo sát (Questionnaire)

3.4. Thu thập dữ liệu

3.5. Tiền xử lý dữ liệu

3.6. Định lượng tập giá trị các biến đặc trưng của không gian năng lực

3.7. Định lượng giá trị các biến đặc trưng phi học tập của vector năng lực của người được khảo sát

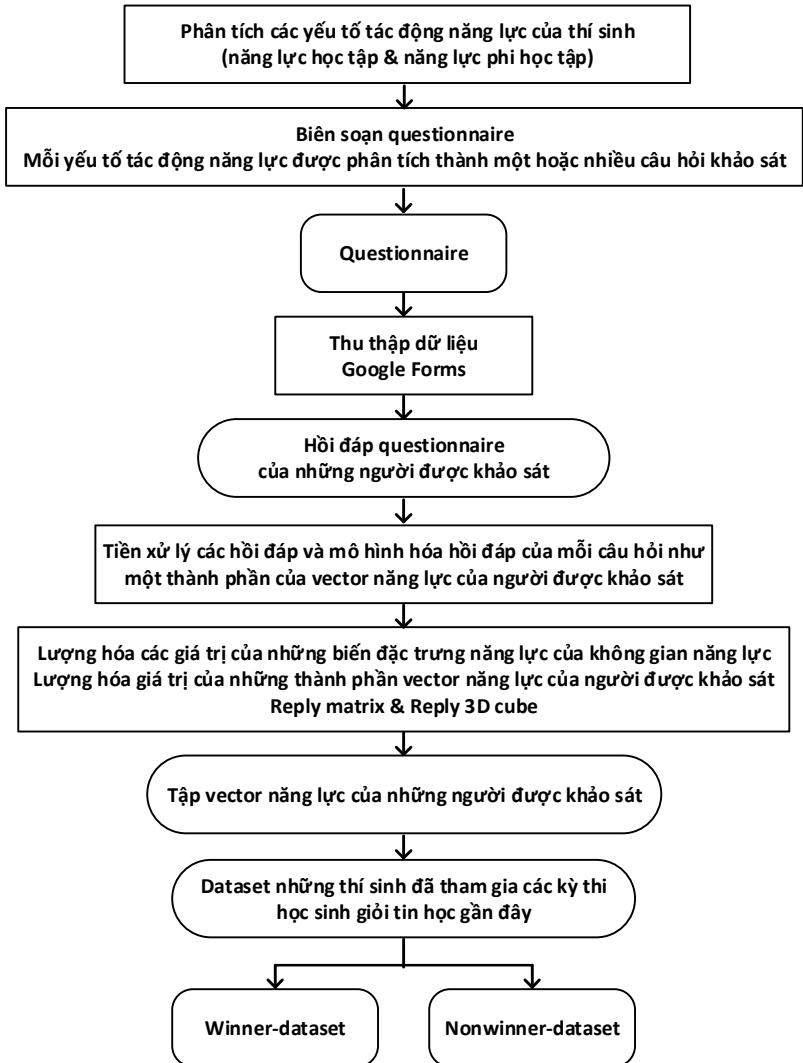
3.8. Dataset

3.9. Kết luận chương 3

Tóm tắt những nội dung chính của chương 3

Dataset sử dụng trong các nghiên cứu về dữ liệu giáo dục cần phải phù hợp với mục tiêu nghiên cứu, môi trường giáo dục, đối tượng nghiên cứu, và người dùng kết quả nghiên cứu. Do đó, các công trình

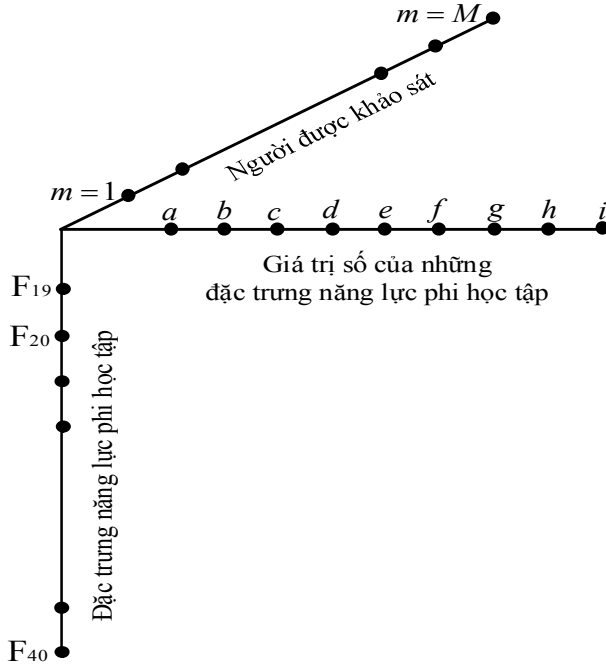
nghiên cứu về dữ liệu giáo dục không dùng chung dataset, mỗi công trình phải thiết kế dataset riêng với số biến dữ liệu, số mẫu phù hợp.



Hình 3.1 Quy trình thiết lập dataset của luận án

Đặc trưng năng lực thí sinh tham gia các kỳ thi học sinh giỏi là những yếu tố năng lực tác động đến kết quả thi, gồm năng lực học tập

và năng lực phi học tập. Luận án đã phân tích các yếu tố năng lực thành những câu hỏi khảo sát đơn giản, trả lời định tính, dễ và nhanh. Năng lực của mỗi người được khảo sát (thí sinh đã tham gia kỳ thi học sinh giỏi) được mô hình hóa như một vector năng lực nhiều biến, mỗi biến tương ứng một hỏi đáp định tính trong bảng khảo sát.



Hình 3. 2. Khối 3D hồi đáp - định lượng các giá trị của các biến đặc trưng năng lực phi học tập $F_{19}, F_{20}, \dots, F_{40}$

Luận án đã thu thập hỏi đáp khảo sát, tiền xử lý, đồng thời đề xuất các tiếp cận ma trận hồi đáp (Bảng 3.5) và khối 3D hồi đáp (Hình 3.2) để lượng hóa các vector năng lực trong không gian năng lực nhiều chiều. Dataset (Bảng 3.7) được thành lập từ kết quả thu thập và xử lý này, với mỗi mẫu đã được gán nhãn winner hoặc non-winner từ người được khảo sát. Quy trình thiết lập dataset được trình bày như hình 3.1.

Bảng 3. 5. Ma trận hỏi đáp về đặc trưng năng lực phi học tập $s=19,20,\dots,40$ của người được khảo sát $m|m=1,2,\dots,M$.

Câu 19	a_{19}^m	b_{19}^m	c_{19}^m	d_{19}^m	e_{19}^m	f_{19}^m			
Câu 20	a_{20}^m	b_{20}^m	c_{20}^m	d_{20}^m					
Câu 21	a_{21}^m	b_{21}^m	c_{21}^m	d_{21}^m	e_{21}^m	f_{21}^m	g_{21}^m	h_{21}^m	
Câu 22	a_{22}^m	b_{22}^m	c_{22}^m	d_{22}^m	e_{22}^m				
Câu 23	a_{23}^m	b_{23}^m	c_{23}^m	d_{23}^m					
Câu 24	a_{24}^m	b_{24}^m	c_{24}^m						
Câu 25	a_{25}^m	b_{25}^m							
Câu 26	a_{26}^m	b_{26}^m	c_{26}^m						
Câu 27	a_{27}^m	b_{27}^m	c_{27}^m	d_{27}^m	e_{27}^m	f_{27}^m	g_{27}^m	h_{27}^m	
Câu 28	a_{28}^m	b_{28}^m	c_{28}^m	d_{28}^m					
Câu 29	a_{29}^m	b_{29}^m							
Câu 30	a_{30}^m	b_{30}^m							
Câu 31	a_{31}^m	b_{31}^m							
Câu 32	a_{32}^m	b_{32}^m	c_{32}^m	d_{32}^m	e_{32}^m	f_{32}^m	g_{32}^m	h_{32}^m	i_{32}^m
Câu 33	a_{33}^m	b_{33}^m	c_{33}^m	d_{33}^m	e_{33}^m	f_{33}^m	g_{33}^m	h_{33}^m	
Câu 34	a_{34}^m	b_{34}^m							
Câu 35	a_{35}^m	b_{35}^m	c_{35}^m	d_{35}^m					
Câu 36	a_{36}^m	b_{36}^m	c_{36}^m						
Câu 37	a_{37}^m	b_{37}^m	c_{37}^m	d_{37}^m					
Câu 38	a_{38}^m	b_{38}^m	c_{38}^m	d_{38}^m					
Câu 39	a_{39}^m	b_{39}^m	c_{39}^m	d_{39}^m	e_{39}^m				
Câu 40	a_{40}^m	b_{40}^m	c_{40}^m	d_{40}^m	e_{40}^m				

Bảng 3. 7. Mô hình dataset

Người được khảo sát m	f_1	..	f_{18}	f_{19}	f_{40}
$m = 1$	f_1^1						f_{40}^1
$m = 2$							
..							
..							
$m = M$	f_1^M						f_{40}^M

Kết luận chương 3

Luận án đã đề xuất tiếp cận xây dựng dataset linh hoạt (smart dataset) phù hợp với yêu cầu thành lập đội tuyển dự thi các giải khác nhau. Tiếp cận dataset linh hoạt cung cấp quy trình thành lập tập dữ liệu năng lực của các thí sinh tham dự các kỳ thi học sinh giỏi. Quy trình có thể được áp dụng để thành lập các đội tuyển theo từng môn học khác nhau, theo từng kỳ thi khác nhau, và theo các cấp thi khác nhau.

CHƯƠNG 4. TIẾP CẬN HỌC MÁY ĐỂ CHỌN HỌC SINH GIỎI VÀO ĐỘI TUYỂN DỰ THI GIẢI HỌC SINH GIỎI

4.1. Giới thiệu

4.2. Những công trình nghiên cứu được liên quan

4.3. Phát biểu bài toán

4.4. Thực nghiệm và đánh giá kết quả thực nghiệm: năm học 2023-2024, 2024-2025,

4.5. Kết luận chương 4

Tóm tắt những nội dung chính của chương 4

Câu hỏi nghiên cứu

Ứng viên nào có nhiều khả năng thắng giải?

Mô tả bài toán

Input:

- Tập những thí sinh đã thắng giải, Winner-dataset $W = \{w_1, \dots, w_N\}$, trong đó mỗi winner w_n được biểu diễn như một vector toán học trong không gian năng lực S chiều.
- Tập những thí sinh đã dự thi nhưng không thắng giải, Nonwinner-dataset $U = \{u_1, \dots, u_M\}$, trong đó mỗi nonwinner u_m được biểu diễn như một vector toán học trong không gian năng lực S chiều.
- Tập những ứng viên $X = \{x_1, \dots, x_L\}$ muốn được chọn vào đội tuyển, trong đó mỗi ứng viên x_l được biểu diễn như một vector toán học trong không gian năng lực S chiều.

Output:

- Tập những ứng viên được quyết định chọn vào đội tuyển $Y_{team} = \{y_1, \dots, y_{K_{team}}\}$

Tiếp cận Winner-domain. Tính chất của Winner-domain (hình 4.1):

- Trọng tâm w_c^p của training set: $w_c^p = [f_{1,c}^p; \dots; f_{s,c}^p; \dots; f_{S,c}^p]$
- Bán kính R^p của training set: $R^p = d_{max}^p(w_c^p, w_i^p), \forall i$

$$\text{với } d_{c,i}^p \equiv d(w_c^p, w_i^p) = \sqrt{(f_{1,c}^p - f_{1,i}^p)^2 + \dots + (f_{S,c}^p - f_{S,i}^p)^2}$$

- Góc hướng θ^p của training set V^p

$$\cos \theta_{c,i}^p = \frac{w_c^p \cdot w_i^p}{|w_c^p| |w_i^p|} = \frac{f_{1,c}^p f_{1,i}^p + \dots + f_{s,c}^p f_{s,i}^p + \dots + f_{S,c}^p f_{S,i}^p}{\sqrt{(f_{1,c}^p)^2 + \dots + f_{S,c}^p)^2} \sqrt{(f_{1,i}^p)^2 + \dots + (f_{S,i}^p)^2}}$$

Các Winner-domains $D_k^p(w_c^p, R_k^p, \theta_k^p)$ được đánh giá theo các hệ số TP, FP, FN của ma trận unnormalized confusion matrix. Những Winner-domains có TP cao, FP thấp, FN thấp không chứa nonwinner

$u_m, \forall u_m \in U | m=1, \dots, M$ là domain lỗi, được chọn làm Winner-domain tốt $Dg^q, u_m \notin Dg^q | m=1, \dots, M | q=1, 2, \dots$ với những tính chất sau:

Bảng 4. 1. Unnormalized confusion matrix

Tập dữ liệu kiểm tra	Kết luận: Thắng giải (cột 1 của matrix)	Kết luận: Không thắng giải (cột 2 của matrix)
Lớp thắng giải (hàng 1 của matrix) $V = \{v_i, i=1, \dots, I\}$	TP : Số mẫu $v_i, v_i \in V$ được đoán ĐÚNG là thắng giải	FN : Số mẫu $v_i, v_i \in V$ được đoán SAI là không thắng giải (bị bỏ sót)
Lớp không thắng giải (hàng 2 của matrix) $U = \{u_j, j=1, \dots, J\}$	FP : Số mẫu $u_j, u_j \in U$ được đoán NHẦM là thắng giải	TN : Số mẫu $u_j, u_j \in U$ được đoán ĐÚNG là không thắng giải

- Trọng tâm w_c^p , cùng với Training set

$$V^p = \{w_1, \dots, w_I\} | p=1, 2, \dots$$

- Bán kính $R_k^p = \alpha_k \cdot R^p$ với $\alpha_k < 1 | k=1, 2, \dots$

- Góc hướng $\theta_k^p \leq \theta^p = (\theta_{c,i}^p)_{\max} | \forall i, i=1, \dots, I_k^p$ với I_k^p là số mẫu của Winner-domain $D_k^p | k=1, 2, \dots$

Độ phức tạp của thuật toán Winner-domain được xác định bởi:

$$O(P.K.(N+M).S)$$

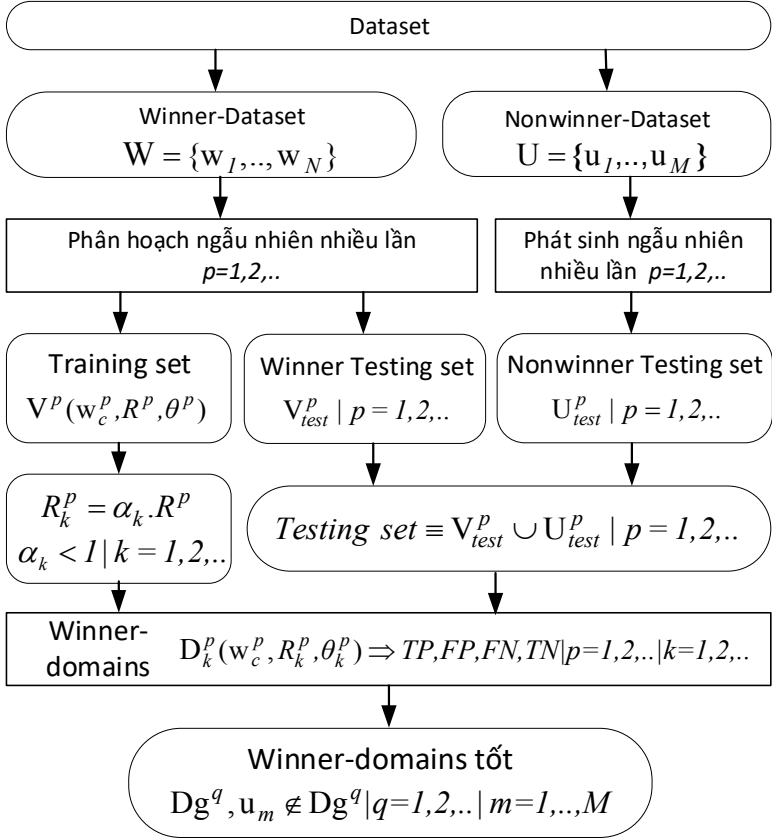
$$T(n) = O(P.K.(N+M).S)$$

Tiếp cận Winner-cosin

Độ tương đồng cosin:

$$\text{CosinSimilarity}(x_l, w_n) \equiv \cos \theta_{ln} = \frac{x_l \cdot w_n}{|x_l| |w_n|} = \frac{f_{1l} \cdot f_{1n} + \dots + f_{sl} \cdot f_{sn} + \dots + f_{sl} \cdot f_{sn}}{\sqrt{(f_{1l})^2 + \dots + f_{sl})^2} \sqrt{(f_{1n})^2 + \dots + (f_{sn})^2}}, \forall n$$

Độ phức tạp tổng thể của thuật toán Winner-cosin:
 $O(L.(S+P)+K.N.S+K.\log K)$ với $K_{team} \ll L$ và $K_{team} \ll N$

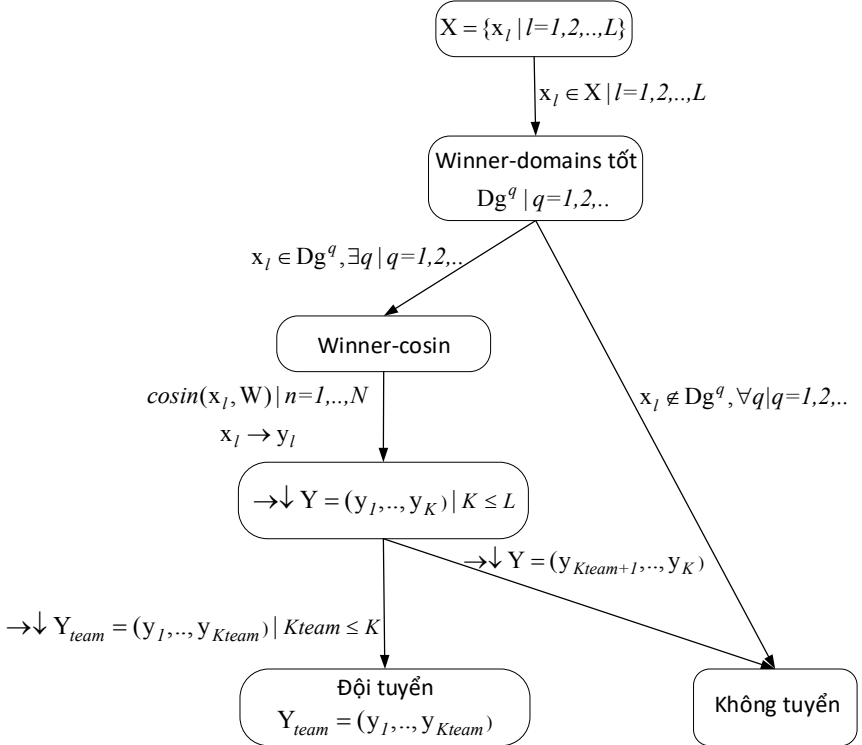


Hình 4. 1. Quy trình thiết lập Winner-domain

Kết luận chương 4

Áp dụng tiếp cận học máy, chương 4 đã xây dựng sơ đồ quyết định chọn những học sinh ứng viên vào đội tuyển trên cơ sở mô hình hóa năng lực mỗi học sinh là một vector nhiều chiều, mỗi chiều biểu diễn một biến đặc trưng năng lực trong không gian năng lực nhiều chiều. Luận án đề xuất các tiếp cận đánh giá độ tương đồng của ứng viên với các winners của vài năm trước gần nhất:

- Tiếp cận Winner-domain đánh giá độ tương đồng về suất năng lực.
- Tiếp cận tương đồng cosin đánh giá độ tương đồng hướng đặc trưng năng lực.



Hình 4. 2. Cây quyết định tuyển chọn đội viên

Với số lượng ứng viên rất nhiều so với số lượng đội viên cần tuyển, luận án không gán nhãn từng ứng viên mà áp dụng chiến lược “thà bỏ lại hơn chọn nhầm” để quyết định chọn ứng viên có khả năng thắng giải cao vào đội tuyển. Nghiên cứu đã được áp dụng thực nghiệm tại một trường phổ thông có kết quả tốt so với những năm trước chưa áp dụng kết quả nghiên cứu của luận án.

Bảng 4.31. Tóm tắt so sánh kết quả thực nghiệm của luận án với kết quả của những phương pháp truyền thống

Năm học	Số lượng đội viên dự thi học sinh giỏi cấp quốc gia môn tin học	Đội viên được tuyển chọn bằng phương pháp truyền thống để dự thi học sinh giỏi cấp quốc gia môn tin học				Đội viên được tuyển chọn bằng mô hình Winner-domain kết hợp Winner-cosin để dự thi học sinh giỏi cấp quốc gia môn tin học				Số lượng đội viên đạt giải học sinh giỏi cấp quốc gia môn tin học	Ghi chú	
		Số lượng	Đạt giải học sinh giỏi cấp quốc gia Số lượng	Tỉ lệ %	Không đạt giải học sinh giỏi cấp quốc gia Số lượng	Tỉ lệ %	Đạt giải học sinh giỏi cấp quốc gia Số lượng	Tỉ lệ %	Không đạt giải học sinh giỏi cấp quốc gia Số lượng			Tỉ lệ %
2019-2020	6	6	0	0%	6	100%					0	
2020-2021	6	6	1	16.67%	5	83.33%					1	
2021-2022	6	6	3	50%	3	50%					3	
2022-2023	6	6	1	16.67%	5	83.33%					1	
2023-2024	10	10	5	50%	5	50%	5	100%	0	0%	5	Mô hình Winner-domain kết hợp Winner-cosin
2024-2025	10	10	6	60%	4	40%	6	100%	0	0%	6	

Thực nghiệm áp dụng cho kỳ thi năm học 2023-2024 với dataset 2024 và tập ứng viên 2024 là những đội viên 2024 đã chọn theo phương pháp truyền thống, kỳ thi năm học 2024-2025 với dataset 2025 và tập ứng viên 2025 là những đội viên 2025 đã chọn theo phương pháp truyền thống. Kết quả được so sánh với 3 mô hình: - mô hình truyền thống, mô hình Winner-domain kết hợp Winner-cosin, mô hình Random Forest theo Bảng 4.32 như sau:

Bảng 4.32. So sánh kết quả các mô hình tuyển chọn và dự đoán kết quả của 2 kỳ thi học sinh giỏi cấp quốc gia môn Tin học năm học 2023-2024 và năm học 2024-2025

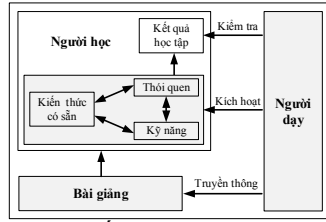
Tiêu chí so sánh	2023-2024		2024-2025	
	Mô hình Winner-domain kết hợp Winner-cosin	Mô hình Random Forest	Mô hình Winner-domain kết hợp Winner-cosin	Mô hình Random Forest
{Đội viên đã được chọn theo mô hình truyền thống} = {Ứng viên input của các mô hình máy tính}	10		10	
Số ứng viên được chọn làm đội viên hoặc dự đoán thắng giải	05	08	06	08
Số đội viên thực tế thắng giải	05	05	06	06
Tỉ lệ chọn đúng hoặc đoán đúng thắng giải: <i>TPR</i>	100%	62.55%	100%	75.0%
Tỉ lệ chọn nhầm hoặc đoán nhầm thắng giải: <i>FPR</i>	0%	0.6%	0%	0.5%
Tỉ lệ bỏ sót: <i>FNR</i>	0%	0%	0%	0%
Tỉ lệ đánh giá hoặc đoán đúng không thắng giải: <i>TNR</i>	1%	0.4%	1%	0.5%

Những kết quả nghiên cứu của luận án đã được áp dụng, có đối sánh với phương pháp truyền thống tuyển chọn và huấn luyện đội môn Tin học của một trường trung học phổ thông với kết quả kỳ thi học sinh giỏi cấp quốc gia các năm học 2023-2024, 2024-2025 (Bảng 4.31). Luận án cũng so sánh kết quả tuyển chọn bằng mô hình Winner-domain kết hợp Winner-cosin với kết quả dự đoán của mô hình Random Forest cho kỳ thi năm học 2023-2024 và 2024-2025 (Bảng 4.32).

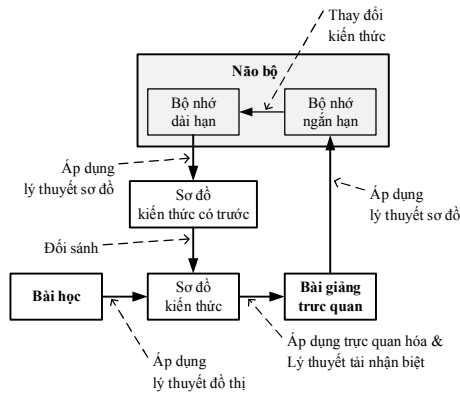
CHƯƠNG 5. TIẾP CẬN HỌC KHÁM PHÁ CÓ HƯỚNG DẪN TRỰC QUAN ĐỂ HUẤN LUYỆN ĐỘI TUYỂN

5.1. Học khám phá có hướng dẫn. Học khám phá có hướng dẫn phát triển từ lý thuyết học khám phá để áp dụng trong các cơ sở giáo dục theo mô hình 3 hợp phần: người học, tài liệu học tập, người dạy.

5.2. Học khám phá có hướng dẫn trực quan. Học khám phá có hướng dẫn trực quan phát triển từ mô hình học khám phá có hướng dẫn với bài giảng trực quan, cấu trúc như hình 5.1, hình 5.2.

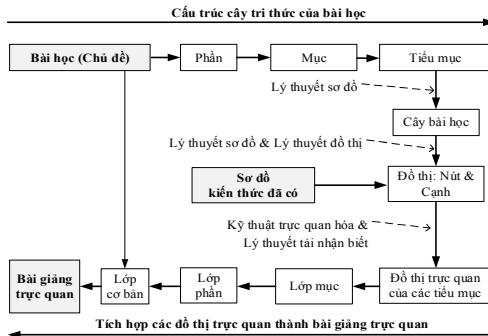


Hình 5. 1. Cấu trúc của hệ thống học khám phá có hướng dẫn trực quan



Hình 5. 2. Dòng thông tin trong quy trình học của người học trong hệ thống học khám phá có hướng dẫn trực quan

5.3. Đánh giá hệ thống học khám phá có hướng dẫn trực quan.



Hình 5. 3. Quy trình thiết kế bài giảng trực quan theo tiếp cận bài học và hướng dẫn theo các bước

Hệ thống học khám phá có hướng dẫn trực quan được đánh giá bằng hiệu năng từng bài giảng trực quan đối với từng nhóm người học và đối với một người dạy bằng cách đối sánh kiến thức người học đạt được so với tải bộ não được huy động và thời gian học.

$$E = \frac{O + K + S + A + M + V + P}{C + T} \times \frac{20}{7} \%$$

E: Hiệu năng của bài giảng trực quan

O: Kiến thức mà người học đã thêm vào bộ nhớ dài hạn sau mỗi bài học.

K: Kiến thức có sẵn của người học

S: Kỹ năng của người học

A: Thói quen của người học

M: Độ tương thích của bài giảng với kiến thức có sẵn của người học

V: Tính trực quan của bài giảng

P: Năng lực sư phạm của người dạy

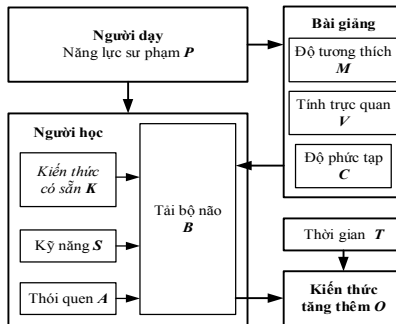
C: Độ phức tạp của bài giảng

T: Thời gian tiếp thu và bổ sung kiến thức trong bộ nhớ dài hạn của người học.

5.4. *Áp dụng phương pháp học khám phá có hướng dẫn trực quan để huấn luyện đội tuyển môn Tin học.* Đội tuyển môn Tin học được rèn luyện để giải các đề thi theo quy trình giải bài toán: (1) Nhận dạng và hiểu bài toán; (2) Hoạch định cách giải; (3) Thực hiện kế hoạch; (4) Kiểm tra kết quả

	Bài toán 1	Bài toán 2	Bài toán 1-1	Bài toán 1
Chuyên đề N				
Chuyên đề N-1				
.....				
Chuyên đề 3				
Chuyên đề 2				
Chuyên đề 1				

Hình 5.4. Quy trình thiết kế bài giảng trực quan theo tiếp cận bài toán



Hình 5.5. Mô hình đánh giá hiệu năng của hệ thống học khám phá có hướng dẫn trực quan áp dụng cho 1 bài giảng.

5.5. *Kết luận chương 5.* Luận án đã đề xuất mô hình học khám phá có hướng dẫn trực quan kết hợp với các tính chất của phương pháp học cộng tác và học theo tiếp cận bài toán. Mô hình học khám phá có hướng dẫn trực quan không chỉ phù hợp sự phát triển của xã hội, sự phát triển của các nguồn dữ liệu, nguồn thông tin, các phương tiện kỹ thuật trên internet và đặc biệt là phù hợp trong việc huấn luyện đội viên đội tuyển môn Tin học.

KẾT LUẬN

Luận án đã thực hiện được những kết quả sau:

1. Dataset. *Luận án đã thiết lập tập dữ liệu (dataset) năng lực thí sinh đã tham gia các kỳ thi học sinh giỏi cấp quốc gia môn Tin học, gồm những thí sinh thắng giải (winners) và những thí sinh không thắng giải (nonwinners).*

Luận án đã xây dựng mô hình vector đặc trưng năng lực của những thí sinh gồm những biến đặc trưng học tập (learning features) và những biến đặc trưng phi học tập (non-learning features). Tiếp cận ma trận hỏi đáp và khối 3D hỏi đáp được đề xuất để định lượng các tập biến đặc trưng năng lực của không gian năng lực và định lượng các biến đặc trưng năng lực của vector năng lực của người được khảo sát. Kết quả là luận án đã xây dựng được 2 datasets: dataset 2024 gồm 221 mẫu (120 winners và 101 nonwinners) của những thí sinh các kỳ thi năm học 2020-2021, 2021-2022, 2022-2023; dataset 2025 gồm 497 mẫu (360 winners và 137 nonwinners) của những thí sinh các kỳ thi năm học 2021-2022, 2022-2023, 2023-2024.

Phương pháp thiết lập dataset của luận án có tính linh hoạt phù hợp tiếp cận Winner-domain, tiếp cận Winner-cosin và phù hợp với đối tượng tạo mẫu là thí sinh tham gia các kỳ thi học sinh giỏi cấp quốc gia. Dataset linh hoạt có thể thay đổi dễ dàng về số mẫu, có thể được cập nhật về số lượng, nội dung của biến dữ liệu theo yêu cầu thành lập đội tuyển của môn thi, cấp thi, kỳ thi.

2. Mô hình Winner-domain kết hợp Winner-cosin. Luận án đã xây dựng mô hình tuyển chọn học sinh giỏi vào đội tuyển học sinh giỏi môn Tin học của một trường trung học phổ thông để tham dự kỳ thi học sinh giỏi cấp quốc gia môn Tin học bằng cách kết hợp các thuật toán Winner-domain và Winner-cosin do luận án đề xuất.

Dựa theo nguyên lý “gần mực thì đen gần đèn thì sáng” và chiến lược “thà bỏ lại hơn chọn nhầm”, luận án đã đề xuất thuật toán Winner-domain để xác định miền thắng giải và thuật toán Winner-cosin cho mô hình cây quyết định để chọn đội viên cho đội tuyển học sinh giỏi môn Tin học từ nhiều ứng viên là học sinh giỏi. Luận án đã được áp dụng thực nghiệm các kỳ thi năm học 2023-2024 và 2024-2025.

3. Mô hình học khám phá có hướng dẫn trực quan. Luận án đã xây dựng mô hình học khám phá có hướng dẫn trực quan.

Luận án đã kết hợp mô hình học khám phá với các mô hình học cộng tác, học theo dự án, ..., và các lý thuyết sơ đồ, lý thuyết tải nhận biết của con người trong quy trình tiếp nhận và lưu trữ kiến thức, và kỹ thuật trực quan để xây dựng quy trình và phương pháp biên soạn bài giảng trực quan áp dụng cho mô hình học khám phá có hướng dẫn trực quan. Luận án cũng đã xây dựng mô hình toán học để đánh giá hiệu năng (efficiency) của từng bài giảng trực quan.

Hướng tới, luận án sẽ được tiếp tục nghiên cứu để cập nhật giá trị biến đặc trưng năng lực thí sinh của những kỳ thi mới nhất và nghiên cứu các biến đặc trưng năng lực cho những môn thi khác, kỳ thi khác, đồng thời so sánh thêm với các mô hình học máy khác.

**DANH MỤC CÁC BÀI BÁO ĐÃ XUẤT BẢN
LIÊN QUAN ĐẾN LUẬN ÁN**

- [CT.1] **Cam Ngoc Thi Huynh**, Hiep Van Nguyen, Phuoc Vinh Tran, Diu Ngoc Thi Ngo, Trung Vinh Tran, and Hong Thi Nguyen, "An Approach to Selecting Students Taking Provincial and National Excellent Student Exams," in Context-Aware Systems and Applications, C. V. Phan and T. D. Nguyen, Eds. no. Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, 2023, vol 475. Springer. pp.156–161 ISBN 978-3-031-28815-9 (Scopus).
- [CT.2] **Cam Ngoc Thi Huynh**, Phuoc Vnh Tran, and Trung Vinh Tran, "The Approach of Winner-domain to Selecting Members for High School Team Participating in National Excellent Student Competition," Mobile Networks and Applications, 2024, Volume 29, Issue 1, pp. 306-313. (SCIE, Q2 / 2024), (Scopus).
- [CT.3] **Cam Ngoc Thi Huynh**, Anh Van Thi Tran, Tha Thi Bui, Hong Thi Nguyen, and Phuoc Vinh Tran, "Applying Guided Discovery Learning to Enhance the Achievement of Information Technology Team," Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering. Context-Aware Systems and Applications, 2024, vol 579, pp. 186–196 Springer, Cham.ISBN: 978-3-031-58877-8 (Scopus).
- [CT.4] **Cam Ngoc Thi Huynh**, Hong Thi Nguyen, Phuoc Vinh Tran, "Efficiency of Visually-Guided Discovery-Learning System for High School", International Conference on Energy, Infrastructure and Environmental Research (EIER 2025), E3S Web of Conferences 626, 04002 (2025), eISSN: 2267-1242. (Scopus).