

**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 - Năm 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:

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

Phản biện 1: TS. Nguyễn Như Sơn

Phản biện 2: PGS.TS. Lê Kim Hùng

Phản biện 3: PGS.TS. Trần Trung Duy

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 08 giờ 00 ngày 23 tháng 4 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

I. MỞ ĐẦU

1. Lý do chọn đề tài

Mỗi năm, Bộ Giáo dục Đào tạo Việt Nam 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.

Công việc thành lập và huấn luyện đội tuyển đang áp dụng tại nhiều trường gồm:

1. Tuyển học sinh giỏi vào các đội tuyển học sinh giỏi dự thi học sinh giỏi cấp quốc gia theo từng môn thi.
2. Ngay sau thành lập được đội tuyển, huấn luyện viên tổ chức huấn luyện đội viên để dự thi.

Với những yêu cầu về thành tích của trường và những thách thức trong công tác tuyển chọn và huấn luyện đội viên đội tuyển, luận án đã đặt ra các nhiệm vụ nghiên cứu:

- 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 môn thi mà đội mình sẽ dự thi?
- 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?

2. Mục tiêu của luận án

Mục tiêu nghiên cứu: 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 môn Tin học với kỳ vọng thắng được nhiều giải, dựa trên tiếp cận học máy và trực quan hóa dữ liệu trong khoa học máy tính.

Nội dung nghiên cứu:

- Phân tích các yếu tố năng lực của những thí sinh đã thắng giải (winners) và những thí sinh đã không thắng giải (nonwinners) trong những kỳ thi trước để mô hình hóa toán học mỗi người như một vector trong không gian nhiều chiều, mỗi chiều biểu diễn một biến 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 và giáo viên huấn luyện đội tuyển.

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 thông tin 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 của 3 năm trước gần nhất.

- Đặc trưng năng lực học sinh gồm những đặc trưng học tập (learning features) và những đặc trưng phi học tập (non-learning features) được phân tích thành những biến năng lực.
- Mô hình học khám phá có hướng dẫn để huấn luyện 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 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 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 tìm ra những đội viên tiềm năng từ những ứng viên dự tuyển là học sinh giỏi của toàn trường.
- *Phương pháp cây quyết định.* Luận án áp dụng phương pháp cây quyết định với thuật toán Winner-domain và Winner-cosin (cosine similarity) để quyết định chọn đội viên đội tuyển từ tập ứng viên.
- *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 để tìm ra đội viên tiềm năng từ những ứng viên là học sinh giỏi của trường và thuật toán Winner-cosin với cây quyết định để chọn đội viên đội tuyển từ những đội viên tiềm năng.

6. Các đóng góp của luận án

6.1. Kết quả nghiên cứu nội dung 1

- Phương pháp biên soạn câu hỏi khảo sát định tính, đơn giản.
- Phương pháp thành lập các biến năng lực và định lượng các hỏi 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

- Thuật toán Winner-domain xác định vùng năng lực của các winners trong không gian năng lực nhiều chiều để tìm ra những đội viên tiềm năng từ những ứng viên là học sinh giỏi của toàn trường,
- Tiếp cận cây quyết định với thuật toán Winner-cosin xác định tương đồng theo hướng năng lực của từng đội viên tiềm năng với tất cả winners trong Winner-Dataset để chọn ra đội viên đội tuyển.

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 bài giảng áp dụng cho mô hình đối với từng bài giảng của thầy hướng dẫn.

7. Cấu trúc luận án

Luận án được thiết kế gồm phần mở đầu, bốn 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 cần nghiên cứu.

Chương 1 trình bày tổng quan các công trình, lý thuyết về khai phá dữ liệu giáo dục và học khám phá được liên quan trong nghiên cứu.

Chương 2 thiết kế dataset là tập năng lực của những thí sinh đã thắng giải (Winners) và những sinh không thắng giải (Nonwinners) của 3 năm liền trước kỳ thi. Dataset được thiết lập theo quy trình: Phân tích các yếu tố đặc trưng năng lực (năng lực học tập và năng lực phi học tập) → Mô hình hóa năng lực như một vector toán học nhiều biến → Thiết kế bộ câu hỏi khảo sát (Questionnaire) → Thu thập hồi đáp định tính → Lượng hóa các hồi đáp → Định lượng các tập biến năng lực của không gian năng lực và định lượng các biến của từng vector năng lực của người được khảo sát → Thiết lập dataset. Luận án đã biên soạn câu hỏi định tính để khảo sát năng lực những thí sinh đã tham dự các kỳ thi học sinh giỏi Tin học cấp quốc gia trong 3 năm liền trước cho 2 kỳ thi học sinh giỏi Tin học cấp quốc gia năm học 2023-2024 và 2024-2025, và những ứng viên dự tuyển với yêu cầu hồi đáp hoàn toàn định tính. Luận án đề xuất tiếp cận ma trận hồi đáp và khối 3D hồi đáp để lượng hóa các biến năng lực và các vector năng lực của người được khảo sát. Với những vector năng lực của những thí sinh dự thi học sinh giỏi Tin học cấp quốc gia của 3 năm liền trước cho 2 kỳ thi học sinh giỏi Tin học cấp quốc gia năm học 2023-2024 và 2024-2025, luận án đã lần lượt thành lập được Dataset 2024 và Dataset 2025, mỗi Dataset gồm Winner-Dataset và Nonwinner-Dataset.

Chương 3 trình bày tiếp cận tuyển chọn đội viên với 2 giai đoạn: Ứng viên dự tuyển → Đội viên tiềm năng → Đội viên đội tuyển. Luận án áp dụng tiếp cận học máy có giám sát để thành lập các Winner-domains và đề xuất tiếp cận Winner-domain tốt từ nhiều Winner-domain để chọn những ứng viên dự tuyển vào đội tiềm năng, và đề xuất tiếp cận Winner-cosin với độ tương đồng cosin cùng với cây quyết định để tuyển chọn đội viên đội tuyển từ đội viên tiềm năng.

Chương 4 trình bày tiếp cận học khám phá và các nguyên lý tiếp nhận thông tin, kiến tạo tri thức của con người để đề xuất mô hình học khám phá có hướng dẫn trực quan áp dụng trong học nhóm và hướng dẫn nhóm để huấn luyện đội tuyển. Luận án cũng xây dựng tiếp cận tính hiệu năng của một bài giảng áp dụng mô hình.

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 LÝ THUYẾT VÀ NGHIÊN CỨU VỀ KHAI PHÁ DỮ LIỆU GIÁO DỤC VÀ HỌC KHÁM PHÁ ĐƯỢC LIÊN QUAN TRONG LUẬN ÁN

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.
- Nguyên lý thu thập thông tin và kiến tạo tri thức của con người; các mô hình, tính chất và phương pháp học khám phá và học khám phá có hướng dẫn.

CHƯƠNG 2. DATASET NĂNG LỰC THÍ SINH THAM GIA CÁC KỲ THI HỌC SINH GIỎI

2.1. Giới thiệu

2.2. Các yếu tố năng lực thí sinh

2.3. Câu hỏi khảo sát (Questionnaire)

2.4. Thu thập dữ liệu

2.5. Tiền xử lý dữ liệu

2.6. Định lượng tập giá trị các biến năng lực của không gian năng lực

2.7. Định lượng giá trị các hợp phần phi học tập của vector năng lực của người được khảo sát.

2.8. Dataset

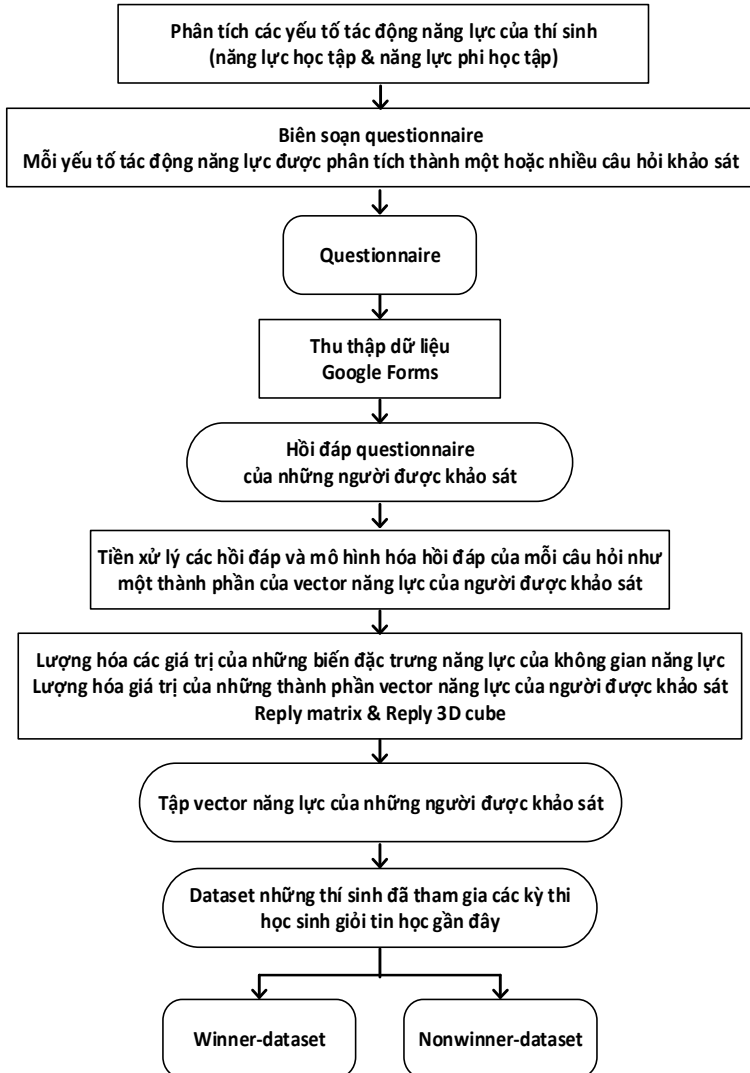
2.9. Kết luận chương 2

Tóm tắt những nội dung chính của chương 2

Khi khảo sát nhiều công bố quốc tế về những nghiên cứu liên quan dữ liệu giáo dục áp dụng EDM (Educational Data Mining), tác giả chưa thấy có công trình sử dụng chung Dataset. Có trường hợp, trong một nghiên cứu có những Datasets được thiết lập riêng cho các bài toán khác nhau. Khảo sát nội dung của những công trình này có thể hiểu rằng Dataset cần phải được thiết lập phù hợp với mục tiêu nghiên cứu, môi trường giáo dục, đối tượng nghiên cứu, thời gian nghiên cứu và người dùng kết quả nghiên cứu. Những Datasets của các nghiên cứu cũng khác nhau về số biến dữ liệu, số mẫu được sử dụng trong các công trình từ vài trăm mẫu đến vài ngàn mẫu, không giống nhau.

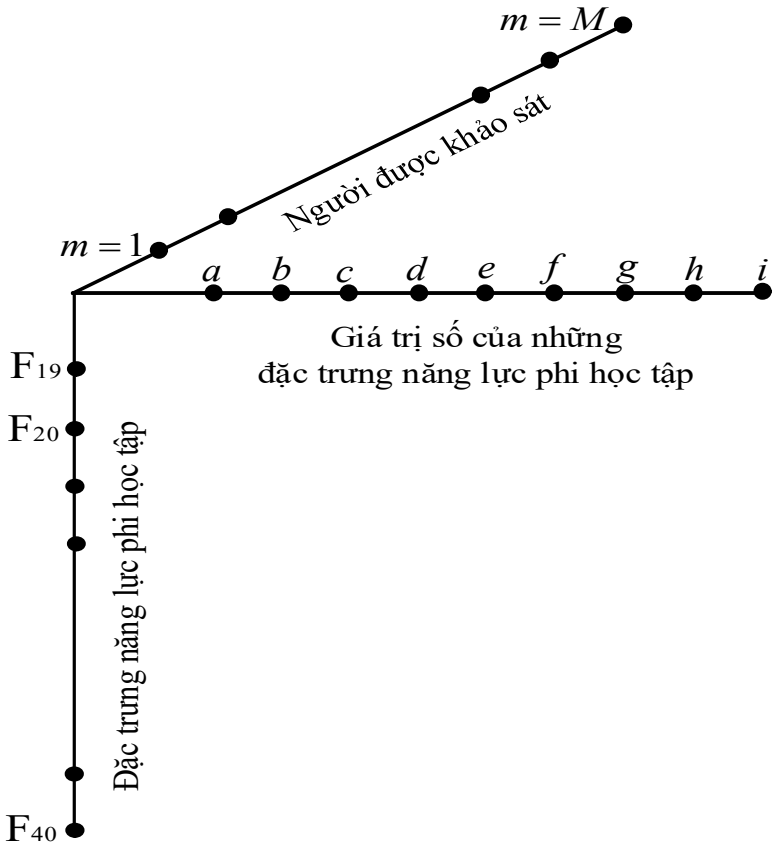
Luận án đã tiến hành nghiên cứu các yếu tố năng lực tác động đến kết quả thi của thí sinh tham gia kỳ thi học sinh giỏi gồm các đặc trưng năng lực học tập và các đặc trưng năng lực phi học tập. Từ đó, luận án đã mô hình hóa mỗi thí sinh, mỗi đội viên như một vector toán học nhiều biến, biến năng lực học tập và biến năng lực phi học tập, trong

không gian năng lực nhiều chiều, mỗi chiều biểu diễn một biến năng lực. Luận án đã phân tích mỗi đặc trưng năng lực thành nhiều câu hỏi khảo sát đơn giản, trả lời định tính, dễ và nhanh, mỗi câu có nhiều đáp án đánh dấu tích (X), được ánh xạ vào mô hình như một biến năng lực.



Hình 2.1 Quy trình thiết lập dataset của luận án

Luận án đã thu thập hồi đáp khảo sát, tiền xử lý, đồng thời đề xuất các tiếp cận ma trận hồi đáp (Bảng 2.5) và khối 3D hồi đáp (Hình 2.2) để lượng hóa các vector năng lực trong không gian năng lực nhiều chiều. Dataset (Bảng 2.7) được thành lập từ kết quả thu thập và xử lý này, với mỗi mẫu đã được gán nhãn sẵn là winner hoặc non-winner từ người được khảo sát. Quy trình thiết lập dataset được trình bày như hình 2.1.



Hình 2.2. Khối 3D hồi đáp - định lượng các giá trị của các biến đặc trưng năng lực phi học tập $F_{19}, F_{20}, \dots, F_{40}$

Bảng 2.5. Ma trận hỏi đáp về 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 2.7. 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 2

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 thành lập 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. Hai Datasets trong luận án này thực hiện mục tiêu tuyển chọn và huấn luyện đội tuyển Tin học để tham dự kỳ thi học sinh giỏi cấp quốc gia năm 2024 và 2025.

CHƯƠNG 3. TIẾP CẬN HỌC MÁY ĐỂ CHỌN HỌC SINH GIỎI VÀO ĐỘI TUYỂN DỰ THI GIẢI HỌC SINH GIỎI QUỐC GIA

3.1. Giới thiệu

3.2. Những công trình nghiên cứu được liên quan

3.3. Phát biểu bài toán

3.4. Thực nghiệm và đánh giá kết quả áp dụng năm học 2023-2024, 2024-2025,

3.5. Kết luận chương 3

Tóm tắt những nội dung chính của chương 3

Câu hỏi nghiên cứu

Ứng viên nào có nhiều khả năng thắng giải ?

Mô tả bài toán

Input:

- Tập winners, gọi là winner-dataset $W = \{w_1, \dots, w_N\}$. Mỗi winner w_n được biểu diễn như một vector toán học trong không gian năng lực S chiều.
- Tập những thí sinh đã tham gia các kỳ thi mà không thắng giải, nonwinner-dataset $U = \{u_1, \dots, u_M\}$. 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 trở thành đội viên đội tuyển. 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 chọn làm đội viên đội tuyển $Z_{team} = \{z_1, \dots, z_{K_{team}}\}$

Phase 1: Áp dụng tiếp cận Winner-domain

Input:

$$W = \{w_1, \dots, w_N\}$$

$$U = \{u_1, \dots, u_M\}$$

$$X = \{x_1, \dots, x_L\}$$

Output:

Tập những đội viên tiềm năng: $Y = \{y_1, \dots, y_K\} \mid K < N$

Solution:

Tiếp cận Winner-domain được thành lập theo qui trình hình

3.1 có các tính chất:

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

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

Những Winner-domains $D_k^p(w_c^p, R_k^p, \theta_k^p)$ với $R_k^p = \alpha_k \cdot R^p$ được đánh giá bằng các hệ số TP , FP , FN của unnormalized confusion matrix (Bảng 3.1). Những Winner-domains có TP cao, FP thấp, FN thấp và không chứa bất kỳ một nonwinner $u_m, \forall u_m \in U | m=1, \dots, M$ nào là những domains lõi, gọi là Good Winner-domain $Dg^q, u_m \notin Dg^q | m=1, \dots, M | q=1, 2, \dots$:

- Trọng tâm w_c^p của tập huấn luyện $V^p = \{w_1, \dots, w_I\} | p=1, 2, \dots$
- Bán kính $R_k^p = \alpha_k R^p$ 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 algorithm được tính:

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

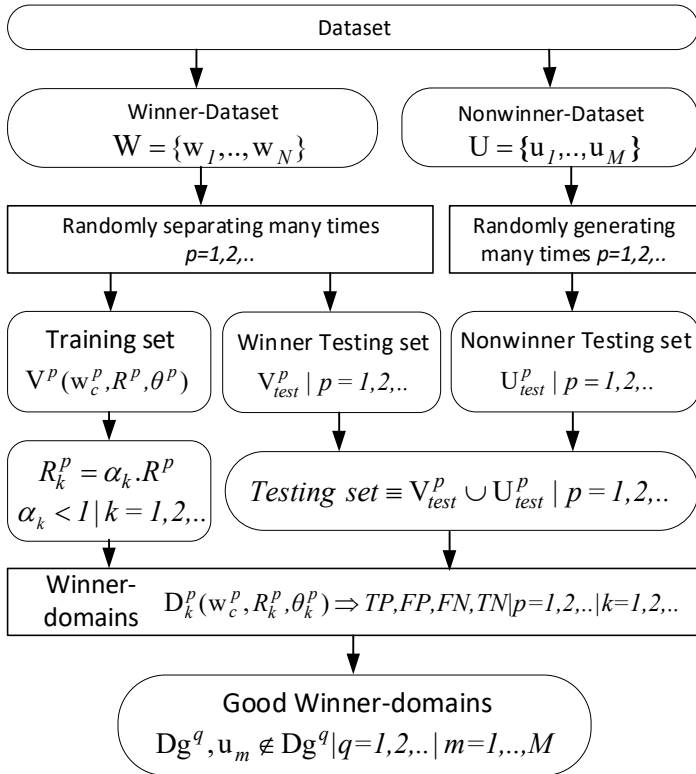
$$T(n)=O(P.K.(N+M).S)$$

Table 3.1. Unnormalized confusion matrix

Testing data set	Conclusion: Win (column 1 of matrix)	Conclusion: Non win (column 2 of matrix)
Win class (line 1 of matrix) $V = \{v_i, i=1, \dots, I\}$	TP : The number of instances $v_i, v_i \in V$	FN : The number of instances $v_i, v_i \in V$ are

	are predicted: TRUE POSITIVE	predicted FALSE NEGATIVE (missed)
Non win class (line 2 of matrix) $U = \{u_j, j=1, \dots, J\}$	FP : The number of instances $u_j, u_j \in U$ are predicted FALSE POSITIVE	TN : The number of instances $u_j, u_j \in U$ are predicted TRUE NEGATIVE

Kết quả của tiếp cận Winner-domain là tập các đội viên tiềm năng
 $Y = \{y_1, \dots, y_K\} \mid K < N$



Hình 3.1. Quy trình thiết lập Winner-domain tốt

Phase 2: Áp dụng tiếp cận Winner-cosin. Tiếp cận Winner-cosin đánh giá độ tương đồng cosin của từng đội viên tiềm năng với Winner-dataset.

Input:

Tập các đội viên tiềm năng $Y = \{y_1, \dots, y_K\} \mid K < N$

Output:

Tập các đội viên được chọn $Z = \{z_1, \dots, z_{Kteam}\} \mid Kteam \leq K$

Solution:

Độ tương đồng cosine của một đội viên tiềm năng với một winner:

$$\begin{aligned} \text{CosinSimilarity}(y_k, w_n) &\equiv \cos \theta_{k,n} = \frac{y_k \cdot w_n}{|y_k| |w_n|} \\ &= \frac{f_{l,k} \cdot f_{l,n} + \dots + f_{s,k} \cdot f_{s,n} + \dots + f_{S,k} \cdot f_{S,n}}{\sqrt{(f_{l,k})^2 + \dots + (f_{s,k})^2} \sqrt{(f_{l,n})^2 + \dots + (f_{S,n})^2}}, \forall k, n \end{aligned}$$

Độ tương đồng cosine của một đội viên tiềm năng với Winner-dataset:

$$\text{CosinSimilarity}(y_k, W) \equiv \cos(y_k, W) = \frac{1}{N} \sum_n \frac{y_k \cdot w_n}{|y_k| |w_n|}, \forall k$$

Thuật toán cây quyết định chọn $Kteam$ đội viên tiềm năng có đội tương đồng cosine với Winner-dataset cao nhất.

- Thuật toán Bubble sort ánh xạ tập $Y = \{y_1, \dots, y_K\}$ vào tập Z :

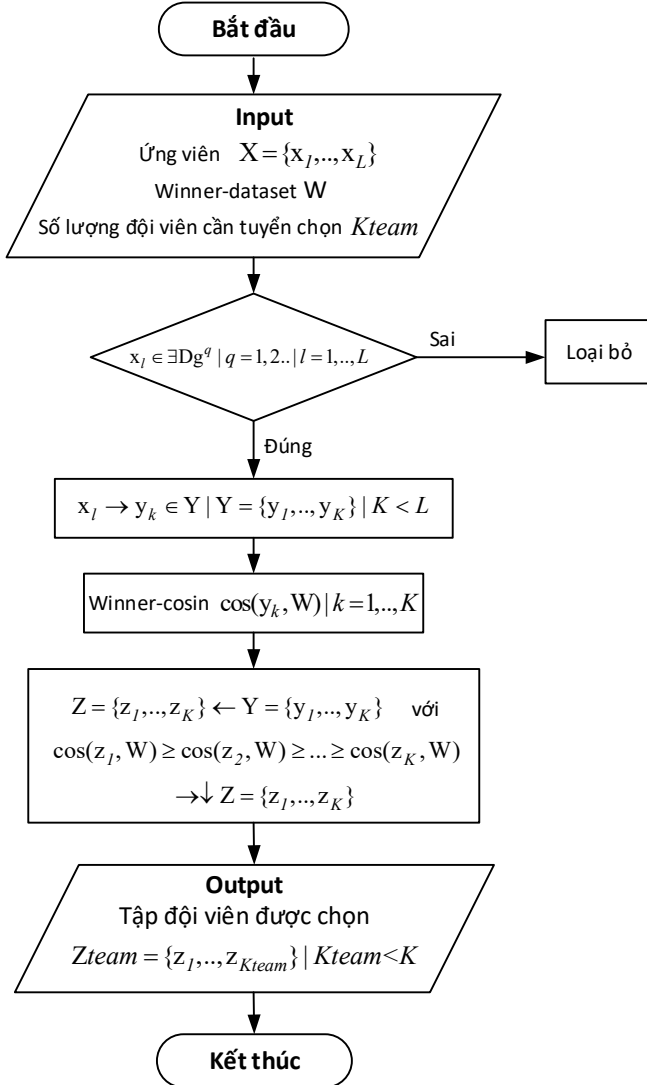
$$Y = \{y_1, \dots, y_K\} \rightarrow Z = \{z_1, \dots, z_K\} \mid \cos(z_1, W) \geq \dots \geq \cos(z_K, W)$$

- Chọn $Zteam = \{z_1, \dots, z_{Kteam}\} \mid Kteam \leq K$

Kết luận chương 3

Áp dụng tiếp cận học máy, chương 3 đã xây dựng sơ đồ quyết định chọn đội viên từ những học sinh ứng viên trên cơ sở mô hình hóa

năng lực mỗi học sinh là một vector nhiều chiều, mỗi chiều biểu diễn một biến 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:



Hình 3.2. Quy trình chọn đội viên từ tập ứng viên

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

Bảng 3.31. So sánh tóm tắt 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
		Đạt giải học sinh giỏi cấp quốc gia			Không đạt giải học sinh giỏi cấp quốc gia			Đạt giải học sinh giỏi cấp quốc gia			Không đạt giải học sinh giỏi cấp quốc gia				
		Số lượng	Tỉ lệ %	Số lượng	Tỉ lệ %	Số lượng	Tỉ lệ %	Số lượng	Tỉ lệ %	Số lượng	Tỉ lệ %				
2019-2020	6	6	0	0%	0	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%	5	100%	0	0%	5	Mô hình Winner-domain kết hợp Winner-cosin	
2024-2025	10	10	6	60%	4	40%	6	100%	6	100%	0	0%	6		

Bảng 3.32. So sánh các mô hình tuyển chọn đội viên bằng 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	
Kỳ thi năm học	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%

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 cao nhất vào đội tuyển vì mục tiêu thắng giải. Nghiên cứu đã được áp dụng thực nghiệm tại một trường trung học phổ thông có kết quả tốt hơn so với những năm trước khi chưa áp dụng kết quả nghiên cứu của luận án.

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

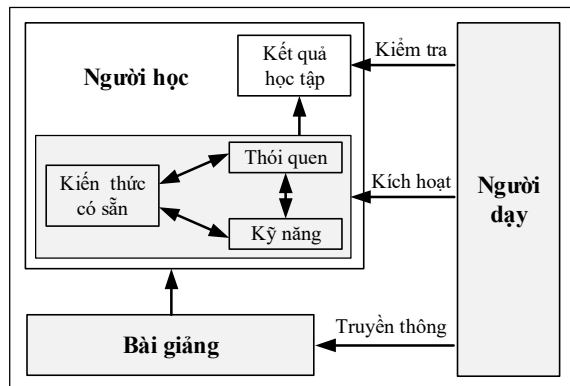
pháp truyền thống. Bảng 3.31 and 3.32 so sánh kết quả của những tiếp cận tuyển chọn đội viên Tin học.

CHƯƠNG 4. TIẾP CẬN HỌC KHÁM PHÁ CÓ HƯỚNG DẪN TRỰC QUAN ĐỂ HUẤN LUYỆN ĐỘI TUYỂN

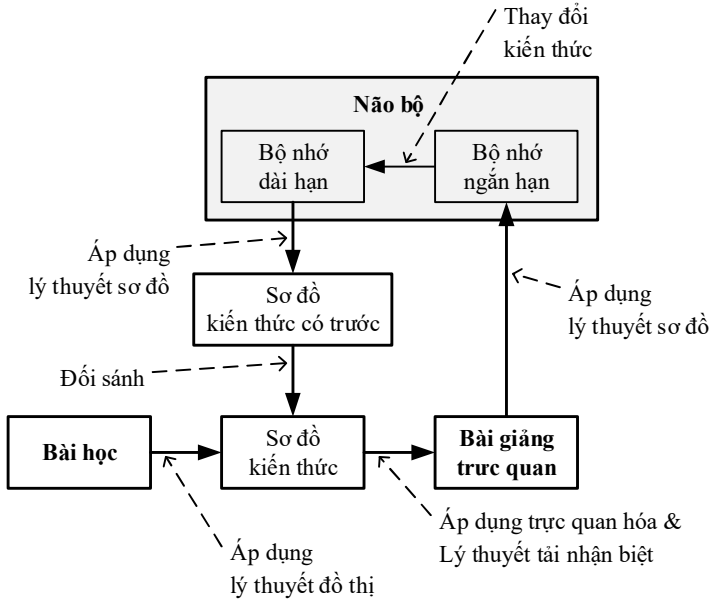
Về phương pháp huấn luyện, Bộ Giáo dục Đào tạo đang đẩy mạnh dự án STEM trong các chương trình phổ thông nhằm hướng tới giảng dạy kết hợp các môn khoa học (science), công nghệ (technology), kỹ thuật (engineering), và toán (mathematic) với phương pháp đặt người học vào trung tâm qui trình dạy và học. 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.

4.1. 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.

4.2. 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 4.1, hình 4.2.

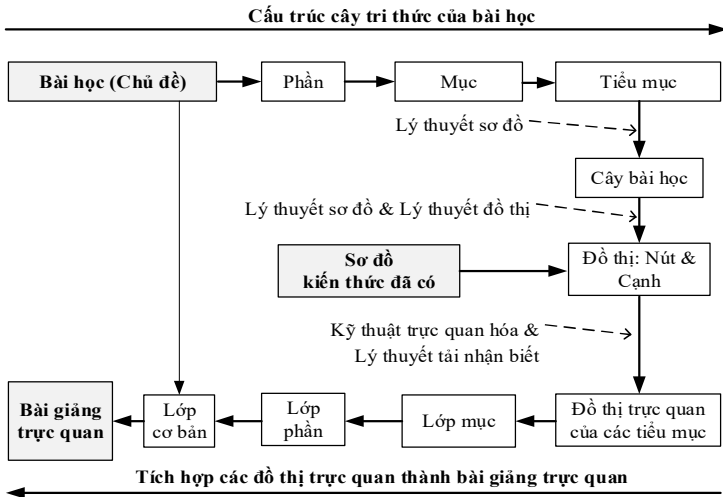


Hình 4.1. Cấu trúc của hệ thống học khám phá có hướng dẫn trực quan



Hình 4.2. Dòng thông tin trong quy trình học của người học trong hệ thống học khám phá có hướng dẫn trực quan

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



Hình 4.3. 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 (hình 4.5).

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

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

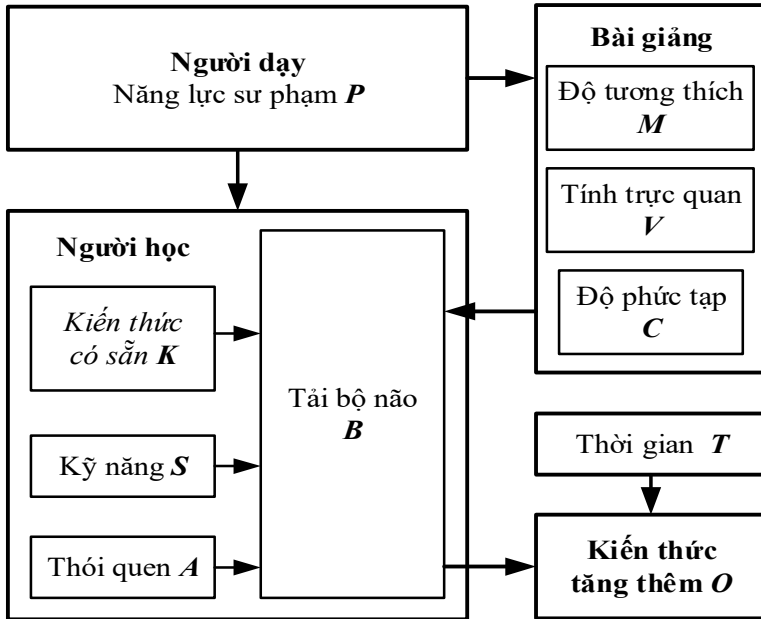
T: Thời gian tiếp thu và bổ sung kiến thức trong bộ nhớ dài hạn của người học.

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

4.4. 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



Hình 4.5. Mô hình đánh giá hiệu năng của hệ thống học khám phá có hướng dẫn trực quan áp dụng cho 1 bài giảng.

Kết luận chương 4

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 (hình 4.4). 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

Kết quả đạt được. Luận án đã thực hiện được những kết quả sau:

1. Dataset. Luận án đã xây dựng qui trình thành lập tập dữ liệu (dataset) năng lực thí sinh 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 mô hình hóa năng lực học sinh như những vector trong không gian năng lực nhiều chiều và đề xuất các tiếp cận ma trận hồi đáp và khối 3D hồi đáp ánh xạ các hồi đáp Questionnaire gồm những câu hỏi có đáp án hoàn toàn định tính thành các biến năng lực và vector năng lực định lượng trong qui trình thiết lập không gian năng lực và biểu diễn năng lực của mỗi người như một vector nhiều chiều định lượng. Dataset được tạo ra bởi luận án chỉ sử dụng cho kỳ thi được thực nghiệm trong luận. Tuy nhiên, phương pháp này có tính linh hoạt, có thể dùng để thiết lập những dataset khác nhau, theo từng môn thi, từng cấp thi, từng kỳ thi bằng cách thay đổi số lượng và nội dung của các biến.

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 của một trường trung học phổ thông để tham dự kỳ thi học sinh giỏi cấp quốc gia bằng cách kết hợp các thuật toán Winner-domain và Winner-cosin do luận án đề xuất.

Áp dụng triết lý “gần mực thì đen, gần đèn thì sáng” và chiến lược “thà bỏ qua hơn chọn nhầm”, luận án đề xuất tiếp cận Winner-domain chọn ra những đội viên tiềm năng từ nhiều ứng viên, tiếp cận Winner-cosin chọn đội viên từ những đội viên tiềm năng có khả năng thắng giải cao nhất với số lượng vừa đủ cho đội. Luận án đã được áp dụng thực nghiệm tại một trường trung học phổ thông để thành lập đội tuyển

các kỳ thi học sinh giỏi cấp quốc gia môn Tin học 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, ..., 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 phát triển của luận án.

- Về phương diện thực tiễn, luận án sẽ được tiếp tục nghiên cứu để cập nhật giá trị biến 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.
- Về phương diện khoa học, luận án sẽ được nghiên cứu để phát triển phương pháp thiết lập tập dữ liệu giáo dục; hoàn thiện quy trình Winner-domain tốt, Winner-cosin và cây quyết định; đánh giá thực tiễn hiệu quả của phương pháp học khám phá có hướng dẫn trực quan áp dụng cho các lớp học.

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).