MINISTRY OF EDUCATION AND TRAINING

VIETNAM ACADEMY OF SCIENCE AND TECHNOLOGY

GRADUATE UNIVERSITY OF SCIENCE AND TECHNOLOGY

HUYNH THI NGOC CAM

APPROACHING MACHINE LEARNING TO SELECTING AND VISUALIZATION TO TRAINING THE MEMBERS OF TEAM

SUMMARY OF DOCTORAL THESIS ON COMPUTER SCIENCE

Code: 9 48 01 01

Ha Noi - 2026

The thesis is completed at: Graduate University of Science and Technology, Vietnam Academy of Science and Technology

Advisor:

Assoc. Prof. Dr. Tran Vinh Phuoc, Ho Chi Minh City University of Natural Resources and Environment

Reviewer 1: Dr. Nguyen Nhu Son

Reviewer 2: Assoc. Prof. Dr. Le Kim Hung

Reviewer 3: Assoc. Prof. Dr. Tran Trung Duy

The thesis was defended with the board assessing doctoral dissertation at Graduate University of Science and Technology, Vietnam Academy of Science and Technology at 08:00, April 23, 2026.

The thesis may be referred at:
1. The library of Graduate University of Science and Technology
2. Vietnam National Library

I. INTRODUCTION

1. Motivation

Annually, the Ministry of Education and Training of Vietnam organizes the National Excellent Student Competition for high school students, with participation in the disciplines such as Informatics, Mathematics, Physics, Chemistry, Biology, etc. To prepare students for this competition, high schools establish teams for each discipline and assign a teacher in charge, known as the coach. Each coach is responsible for selecting teamers (team members) and providing intensive training according to the requirements of the examination before the students participate in the competition.

The process of establishing and training the team, as applied in many schools, includes the following steps:

1. In the light of student's outcome, the coach selects outstanding students for the Informatics team to participate in the National Excellent Student Competition on Informatics.
2. Immediately after forming the team, the coach organizes training for the teamers to prepare for the competition.

Given the school's performance requirements and the challenges in selecting and training teamers, the thesis raises the following research questions:

- How can we identify and understand the competency factors required of a candidate participating in the National Excellent Student Competition for the discipline which the team will compete in?
- How can we select the right high-performing students whose abilities match the requirements of the National Excellent Student Competition?
- Which methods should be applied to train and enhance the competencies of teamers so that they meet the requirements

of the National Excellent Student Competition?

2. Objectives of the thesis

Objective of the research: In the light of the approaches of machine learning and data visualization in computer science, the thesis develops a process for selecting teamers and training them to participate in the National Excellent Student Competition in Informatics, with the expectation winning multiple awards.

Research contents:

- Analyze competency factors of prize-winning contestants (winners) and non-winning contestants (non-winners) in previous competitions, mathematically model each individual as a vector in a multidimensional space, where each dimension represents a competency variable.
- Develop algorithms and procedures to select outstanding students for a high school informatics team to participate in the National Excellent Student Competition on Informatics.
- Build a guided-discovery learning model using visualization techniques, with efficiency evaluation of each lecture for the team and the teachers responsible for training them.

3. Research objects of the thesis

- Competencies of students of high schools who have participated in the National Excellent Student Competition on Informatics.
- Educational data on student competencies.
- Machine learning approaches for educational data mining (EDM: Educational Data Mining).
- Learning theories and discovery learning models (DL: Discovery Learning).
- Principles of human (student) information acquisition and knowledge construction.

4. Scope of the thesis:

- The research data covers prize-winning and non-winning contestants in the National Excellent Student Competition on Informatics over the most recent three years.
- Student competency features include both learning features and non-learning features, which are analyzed and represented as competency variables.
- A visually-guided discovery learning model is developed to train a high school informatics team.

5. Research methodology

- *Analysis–synthesis method:* The analytical method is applied to examine the competency of each student as a multivariate competency vector. The analysis–synthesis approach is also used to structure instructional materials according to a visually-guided discovery learning model.
- *Mathematical modeling method:* Mathematical modeling is employed to represent the competency of each student as a vector in a multidimensional competency space.
- *Machine learning method:* The thesis applies supervised learning methods and proposes Winner-domain algorithm to identify potential teamers from candidates who are high-achieving students across the school.
- *Decision tree method:* The thesis utilizes decision tree in combination with the Winner-domain and Winner-cosin (cosine similarity) algorithms to select teamers from the set of candidates.
- *Mathematical method:* Mathematical methods are applied to construct the dataset.
- *Programming method:* Python software is used to develop and implement the programs for the study.

- *Survey method:* The survey method is applied to design questionnaire, and the web of Google Forms is used to collect data.
- *Experimental method:* Experimental methods are employed to evaluate the Winner-domain algorithm for identifying potential candidates from high-achieving students, and the Winner-cosin algorithm with decision trees for selecting official teamers from the potential pool.

6. Contributions of the Thesis

6.1. The results of the content 1:

- The approach to designing simply qualitative survey questions easy to answer.
- The approach to forming competency variables and quantifying qualitative survey replies and the procedure to built dataset.
- The dataset of contestants who have participated in recent national excellent student competitions on informatics.

6.2. The results of the content 2:

- The approach of Winner-domain defines the competency domain of winners in multi-dimensional competency space to find out potential teamers from several candidates who are outstanding students of the school.
- The approach of decision tree with Winner-cosin algorithm determines the similarity of each potential teamer with all winners in Winner-dataset to select official teamers.

6.3. The results of the content 3:

- The thesis formed a visually-guided-discovery learning model and proposed the approach to evaluating the efficiency of a lecture guided by an instructor.

7. Structure of Thesis

The thesis is designed with an introduction, four content chapters, and a conclusion. The list of the author's published works, the list of references, and the experimental appendix are edited at the end of the thesis. The contents of works are summarized as follow:

- *Introduction:* The thesis introduces the context leading to the topic with 3 challenges to research.
- *Chapter 1:* The chapter presents the overview of related works on educational data mining and discovery learning.
- *Chapter 2:* The chapter designs the contestants' competency dataset, including winners and non-winners, according to the process: Analysing competency feature factors (learning competency factors and non-learning competency factors) → Modeling competency of each student as a mathematical vector of several variables → Designing questionnaire to survey → Collecting qualitative replies → Quantifying qualitative replies → Quantifying competency variable sets of the competency space and quantifying variables of the competency vector of each surveyed individual → Setting up dataset. The thesis edited qualitative questionnaire to survey the competency of contestants who participated in the National Excellent Student Competition on Informatics within 3 recent years for the competitions in academic year 2023-2024 and 2024-2025, and candidates with the requirement for qualitative reply. The thesis proposed the approaches of reply matrix, reply 3D cube to quantify competency variables from the qualitative replies of surveyed individuals. The thesis set up Dataset 2024 and Dataset 2025, each including Winner-Dataset and Nonwinner-Dataset, from the contestants'competency vectors within 3 recent years of 2023-

2024 and 2024-2025.

- *Chapter 3:* The chapter presents the approach to selecting teamers with 2 phases: Candidates → Potential teamers → Teamers. The thesis applies the approach of supervised machine learning to establish Winner-domains and proposes Good Winner-domains approach from several Winner-domains to selecting potential teamers from candidates, and the approach of Winner-cosin along with decision tree, titled Winner-cosine approach, to selecting teamers from potential teamers.
- *Chapter 4:* The chapter proposed the visually-guided-discovery learning model in the light of the human principle of information acquisition and new knowledge development to train a group of students as teamers, and the approach to evaluating the efficiency of a lecture guided by an instructor.
- *Conclusion:* The section summarizes the main contributions of the thesis and plans in future.

II. THESIS CONTENT

CHAPTER 1. OVERVIEW OF RELATED WORKS ON EDUCATIONAL DATA MINING AND DISCOVERY LEARNING

The related works are the researches on students' competency data, algorithms of separating objects into groups by their features. The related works are composed of the following topics:

- Educational data.
- Educational data mining approaches as support vector machine (SVM), decision tree, K-mean clustering, data processing.

- The principle of human information acquisition and knowledge development; the models, characteristics and methods of discovery learning and guided- discovery learning

CHAPTER 2. DATASET OF CONTESTANTS' COMPETENCY

2.1. Introduction
2.2. Competency Factors of Contestants
2.3. Questionnaire
2.4. Data collection
2.5. Data Pre-processing
2.6. Quantifying the value sets of competency variables in competency space.
2.7. Quantifying the non-learning components of contestants' competency vectors.
2.8. Dataset.
2.9. Conclusion of chapter 2

Summary of Main Contents of Chapter 2

The shared dataset in researches on Educational Data Mining has not been found in published works. The dataset of educational data for research on education have to be suitable for research objectives and objects, educational environment, and user of results. Accordingly, the researches relating educational data do not share dataset, each research must design a special dataset of the suitable number of variables and the suitable number of instances, from a few hundred instances to a few thousand instances

Competency features of the contestants participating in national excellent student competitions are the factors impacting on the results, including learning and non-learning competency. The thesis analyzed competency factors into simple, easy and quick-to-answer qualitative

survey questions. The competency of each surveyed person is modeled as a multi-dimensional competency vector, each dimension represents a variable corresponding to a qualitative response in questionnaire.

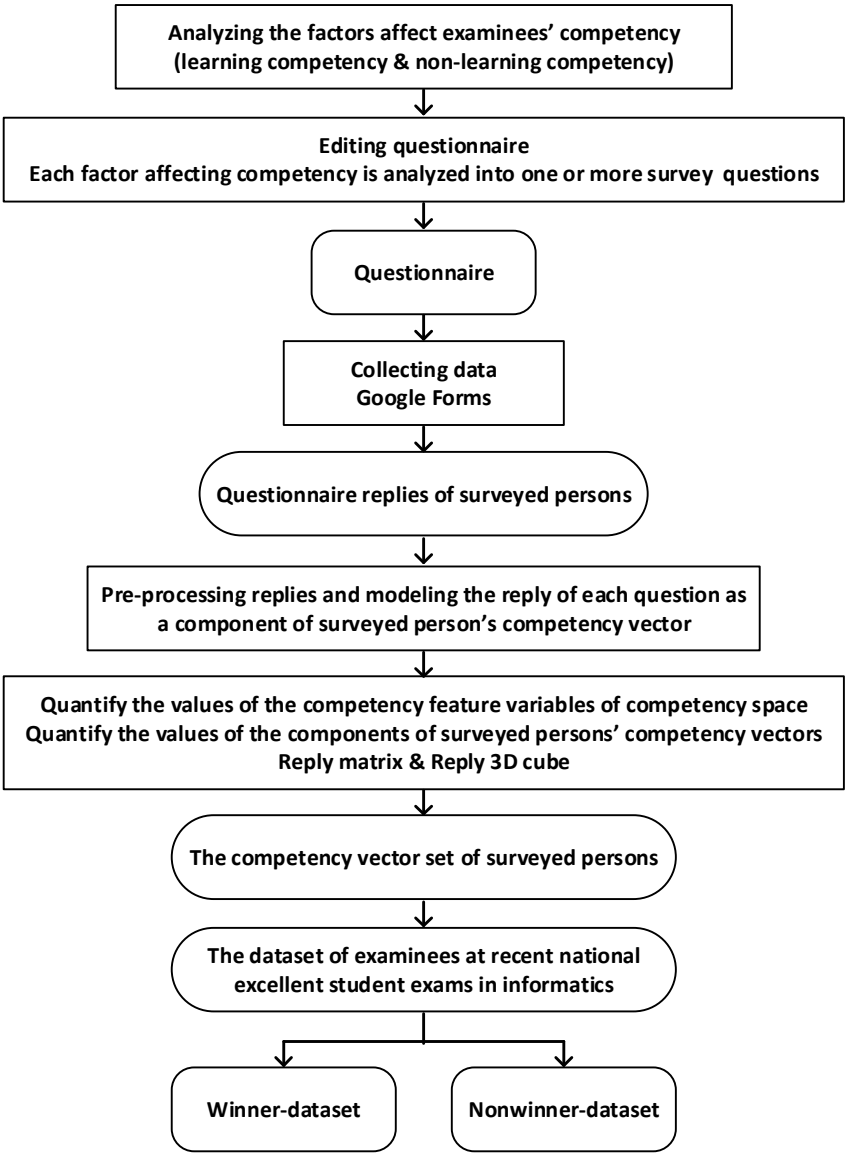

Figure 2.1. The process setting up the dataset of the thesis

The thesis collected and pre-processed replies, proposed the approaches of reply matrix (Table 2.5) and reply 3D cube (Figure 2.2) to quantify the competency vectors in the multi-dimensional competency space. The dataset is set up from the processed replies (Table 2.7), where each instance is labelled with winner or non-winner. The process setting up dataset is presented at the figure 2.1.

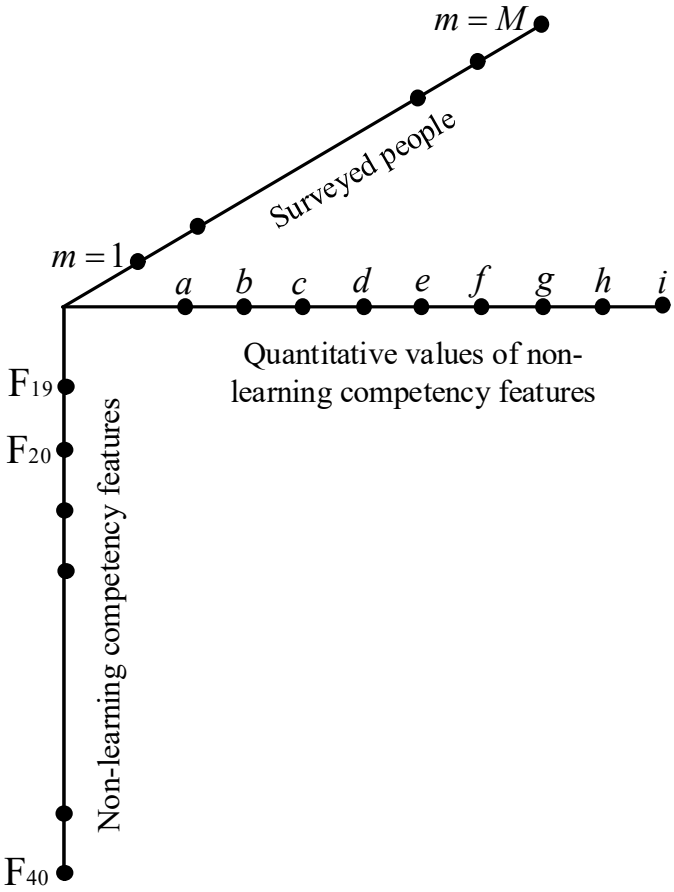

Figure 2.2. Reply 3D cube to quantify the values of non-learning feature variables $F_{19}, F_{20}, .., F_{40}$.

Table 2.5. Reply matrix of the features of non-learning competency s=19,20,..,40 of surveyed person m| m=1,2,..,M.

Question 19	a_{19}^m	b_{19}^m	c_{19}^m	d_{19}^m	e_{19}^m	f_{19}^m			
Question 20	a_{20}^m	b_{20}^m	c_{20}^m	d_{20}^m					
Question 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	
Question 22	a_{22}^m	b_{22}^m	c_{22}^m	d_{22}^m	e_{22}^m				
Question 23	a_{23}^m	b_{23}^m	c_{23}^m	d_{23}^m					
Question 24	a_{24}^m	b_{24}^m	c_{24}^m						
Question 25	a_{25}^m	b_{25}^m							
Question 26	a_{26}^m	b_{26}^m	c_{26}^m						
Question 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	
Question 28	a_{28}^m	b_{28}^m	c_{28}^m	d_{28}^m					
Question 29	a_{29}^m	b_{29}^m							
Question 30	a_{30}^m	b_{30}^m							
Question 31	a_{31}^m	b_{31}^m							
Question 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
Question 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	
Question 34	a_{34}^m	b_{34}^m							
Question 35	a_{35}^m	b_{35}^m	c_{35}^m	d_{35}^m					
Question 36	a_{36}^m	b_{36}^m	c_{36}^m						
Question 37	a_{37}^m	b_{37}^m	c_{37}^m	d_{37}^m					
Question 38	a_{38}^m	b_{38}^m	c_{38}^m	d_{38}^m					
Question 39	a_{39}^m	b_{39}^m	c_{39}^m	d_{39}^m	e_{39}^m				
Question 40	a_{40}^m	b_{40}^m	c_{40}^m	d_{40}^m	e_{40}^m				

Table 2.7. Dataset model

Surveyed people	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

Conclusion of the chapter 2

The thesis proposed the approach to setting up smart dataset suitable for forming the team participating in excellent student competition of various awards. The smart dataset approach enables the process setting up the datasets on the competency of contestants participating in excellent student competitions on various subjects. The process may be applied to form teams for various disciplines. Two datasets in this thesis are formed to carry out the selection and training Informatics team participating in the national excellent student competition in 2024 and 2025.

CHAPTER 3. MACHINE LEARNING APPROACH TO SELECTING MEMBERS FOR TEAMS PARTICIPATING IN NATIONAL EXCELLENT STUDENT COMPETITION

3.1. Introduction

3.2. Related works

3.3. Problem description

3.4. Experiment and estimation of the results of the applications in school years 2023-2024, 2024-2025.

3.5. Conclusion of the chapter

Summary of Main Contents of the Chapter 3

Research Question

Which candidate is likely to win award?

Problem

Input:

- The set of winners, winner-dataset $W=\{w_1,..,w_N\}$. Each winner w_n is represented as a mathematical vector in competency space of S dimensions.
- The set of contestants who participated in competitions but not win, nonwinner-dataset $U=\{u_1,..,u_M\}$. Each nonwinner u_m is represented as a mathematical vector in competency space of S dimensions.
- The set of candidates $X=\{x_1,..,x_L\}$ who want to become teamer. Each candidate x_l is represented as a mathematical vector in competency space of S dimensions.

Output:

- The set of the candidates are selected as teamers $Zteam=\{z_1,..,z_{Kteam}\}$

Phase 1: Applying Winner-domain approach

Input:

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

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

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

Output:

The set of potential teams: $Y=\{y_1,..,y_K\} \mid K < N$

Solution:

The characteristics of winner-domain (Figure 3.1):

- Centroid w_c^p of training set: $w_c^p = [f_{1.c}^p;..;f_{s.c}^p;..;f_{S.c}^p]$
- Radius R^p of training set: $R^p = d_{max}^p(w_c^p, w_i^p), \forall i$

  where $d_{c.i}^p \equiv d(w_c^p, w_i^p) = \sqrt{(f_{1.c} - f_{1.i})^2 + .. + (f_{S.c} - f_{S.i})^2}$
- Directional angle θ^p of training set V^p:

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

The Winner-domains $D_k^p(w_c^p, R_k^p, \theta_k^p)$ with $R_k^p = \alpha_k.R^p$ are evaluated by the coefficients *TP, FP, FN* of the unnormalized confusion matrix (Table 3.1). The Winner-domains with high *TP*, low *FP*, low *FN*, and not including any nonwinner $u_m, \forall u_m \in U \mid m=1,..,M$ are convex domains, titled Good Winner-domain $Dg^q, u_m \notin Dg^q \mid m=1,..,M \mid q=1,2,..$:

- Centroid w_c^p of Training set $V^p = \{w_1,..,w_I\} \mid p=1,2,..$
- Radius $R_k^p = \alpha_k.R^p$ with $\alpha_k < 1 \mid k = 1,2,..$
- Directional angle $\theta_k^p \le \theta^p = (\theta_{c.i}^p)_{\max} \mid \forall i, i = 1,..,I_k^p$ with I_k^p is the number of instances of the Winner-domain $D_k^p \mid k = 1,2,..$

The complexity of Winner-domain algorithm is calculated by:

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

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

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) $V=\{v_i, i=1,..,I\}$	TP : The number of instances $v_i, v_i \in V$ are predicted: TRUE POSITIVE	FN: The number of instances $v_i, v_i \in V$ are predicted FALSE NEGATIVE (missed)
Non win class (line 2 of matrix) $U=\{u_j, j=1,..,J\}$	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

The output of Winner-domain approach is the set of potential teamers $Y=\{y_1,..,y_K\} \mid K < N$

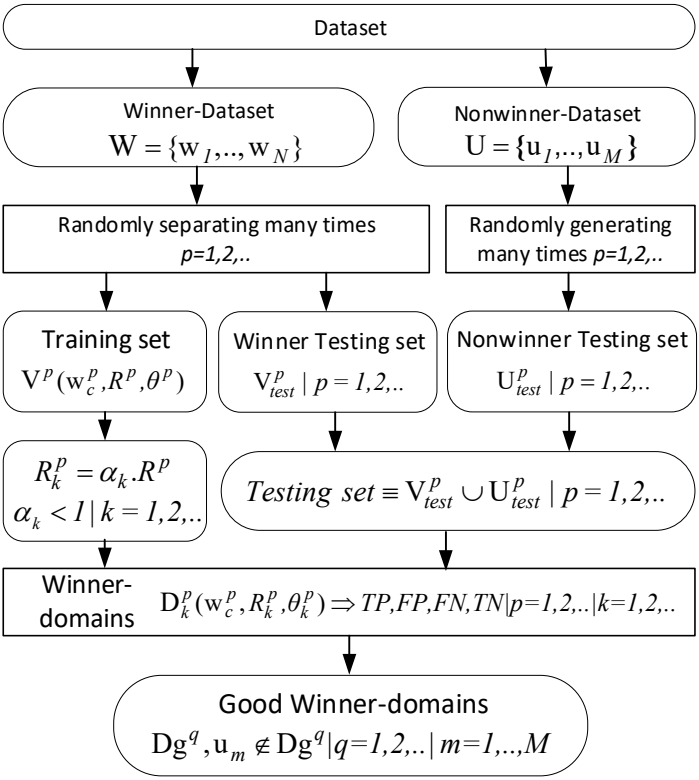

Figure 3.1. Process to form Good Winner-domains

Phase 2: Applying Winner-cosin approach. The approach evaluates the cosin-similarity of each potential teamer with the Winner-dataset.

Input:

The set of potential teamers $Y=\{y_1,..,y_K\} \mid K < N$

Output:

The set of the teamers selected for the team

$Z=\{z_1,..,z_{Kteam}\} \mid Kteam \le K$

Solution:

The cosin-similarity of a potential teamer with a winner:

$$\text{CosinSimilarity}(y_k, w_n) \equiv \cos\theta_{k.n} = \frac{y_k.w_n}{|y_k|\,|w_n|}$$

$$= \frac{f_{1.k}.f_{1.n} + .. + f_{s.k}.f_{s.n} + .. + f_{S.k}.f_{S.n}}{\sqrt{(f_{1.k})^2 + .. + f_{S.k})^2}\sqrt{(f_{1.n})^2 + .. + (f_{S.n})^2}}, \forall k,n$$

The cosin-similarity of a potential teamer with Winner-dataset:

$$\text{CosinSimilarity}(y_k, W) \equiv \cos(y_k, W) = \frac{1}{N}\sum_{n}^{N}\frac{y_k.w_n}{|y_k|\,|w_n|}, \forall k$$

The tree decision selects *Kteam* potential teamers who have the highest Cosin-Similarity with Winner-dataset.

- The Bubble sort algorithm maps the set $Y=\{y_1,..,y_K\}$ into Z:

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

- Select $Zteam=\{z_1,..,z_{Kteam}\} \mid Kteam \le K$

Conclusion of the chapter 3

Applying machine learning, the chapter 3 formed the schema of decision selecting teamers from candidates based on the model of student's competency as multidimensional vector. The thesis proposes the approaches to evaluating the similarity of the candidates with winners of few recent competitions:

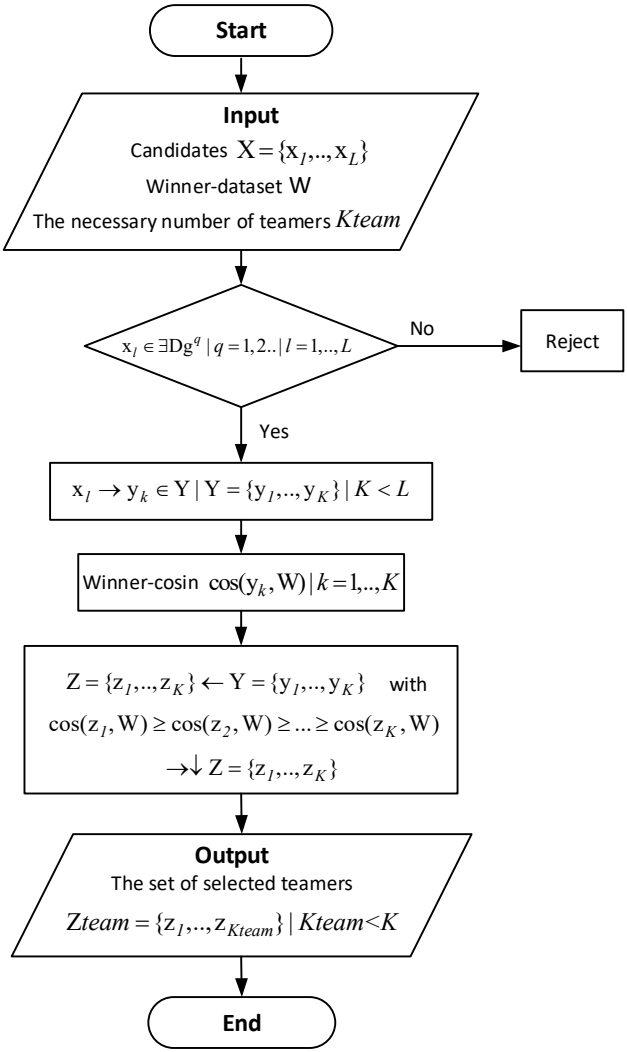

Figure 3.2. Process selecting teamers from the set of candidates

- Winner-domain approach evaluates the similarity of competency magnitudes.
- Winner-cosin approach with cosin-similarity algorithm evaluates the similarity of the competency trend.

Table 3.31. The method of the thesis is experimentally compared with the traditional method.

School years	The number of teamers participating in national excellent student exams in informatics	Teamers are selected with traditional method					Teamers are selected with the model of Winner-domain combining with Winner-cosin						The number of winners	Note
		Quantity (qt)	Winner		Nonwinner		Quantity (qt)	Winner		Nonwinner				
			qt	Ratio %	qt	Ratio %		qt	Ratio%	qt	Ratio %			
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	5	100%	0	0%		5	Model of Winner-domain combining with Winner-cosin
2024-2025	10	10	6	60%	4	40%	6	6	100%	0	0%		6	

Table 3.32. Comparing the models selecting teamers by the results at the 2023-2024 and 2024-2025 national excellent student competitions in informatics.

Criteria for comparing	**2023-2024**		**2024-2025**	
Exams	The model of Winner-domain combining with Winner-cosin	The model of Random Forest	The model of Winner-domain combining with Winner-cosin	The model of Random Forest
{Teamers were selected by traditional model} = {Candidates to input into the models of computer science}	10		10	
The number of candidates are selected as teamers or predicted as winners.	05	08	06	08
The number of teamers really win awards	05		06	
The ratio of true selection of winners or true prediction of winners: *TPR*	100%	62.55 %	100%	75.0 %
The ratio of false selection of winners or false prediction of winners: *FPR*	0%	0.6%	0%	0.5%
The ratio of false nonwinner: *FNR*	0%	0%	0%	0%
The ratio of true estimation or true prediction of nonwinner: *TNR*	1%	0.4%	1%	0.5%

Because of the number of candidates is very big in comparison with the number of necessary teamer, the thesis does not label each candidate. The thesis applies the strategy of “Better to exclude than wrongly include” to select the candidates more likely to win award for team. The work was experimentally applied at a high school with good results in comparison with the years in the past.

The experiment was applied for the 2023-2024 competition with the dataset 2024 and the 2024 candidates being teamers selected by traditional method; the 2024-2025 competition with the dataset 2025 and the 2025 candidates being teamers selected by traditional method. The tables 3.31 and 3.32 compare the results of the approaches selecting Informations teamers.

CHAPTER 4. APPROACHING VISUALLY-GUIDED-DISCOVERY LEARNING TO TRAINING TEAM

4.1. Guided-discovery-learning. Guided-discovery-learning is developed from the theory of discovery learning to apply for educational institutions with the model of 3 components: learners, learning documents, instructors.

4.2. Visually-guided-discovery-learning. Visually-guided-discovery-learning is developed from the model of guided-discovery-learning with visual lecture, structured as the figure 5.1 and 5.2.

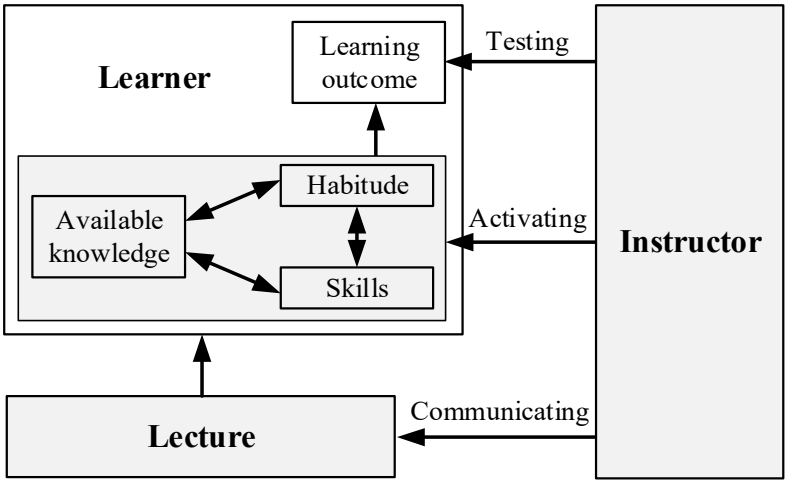

Figure 4.1. The structure of a visually-guided-discovery-learning system.

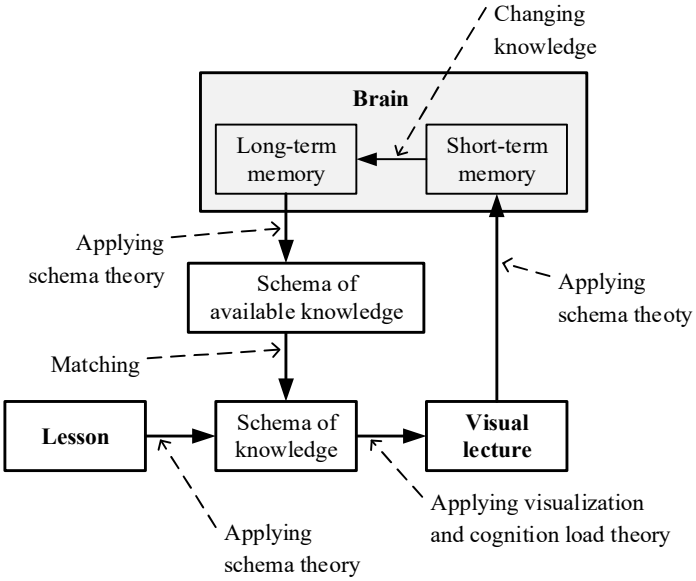

Figure 4.2. The information flow of learners' learning process in visually-guided-discovery-learning system.

4.3. Evaluating visually-guided-discovery-learning system

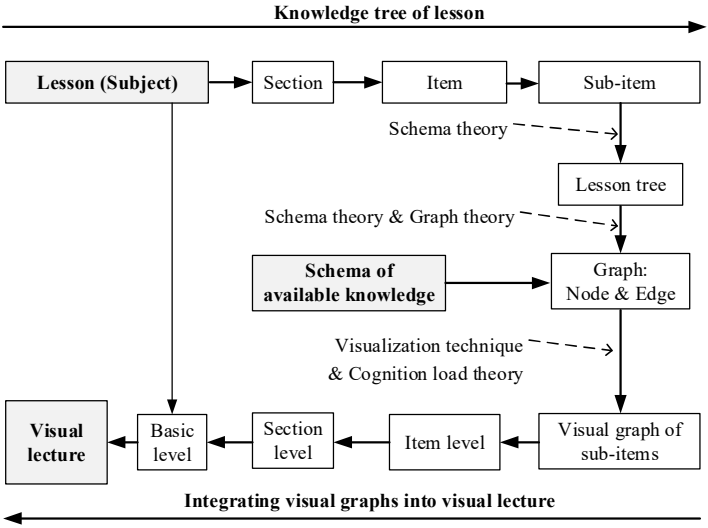

Figure 4.3. The process to design visual lecture according to the lesson.

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

E: The efficiency of visual lecture
O: The knowledge which the learner added to long-term memory after each lecture.
K: The available knowledge of the learner.
S: The skills of the learner
A: The habit of the learner.
M: The matching of lecture with learner' available knowledge.
V: The visual characteristics of the lecture.
P: The pedagogic competence of the instructor
C: The complexity of the lesson.
T: The time necessary to acquire and add new knowledge to the learner's long-term memory.

4.4. Applying the method of visually-guided-discovery learning to train an informatics team. The informatics team is trained to solve exam questions according to the process: (1) Identifying and understanding the questions; (2) Planning how to solve; (3) Developing the plan; (4) Checking results.

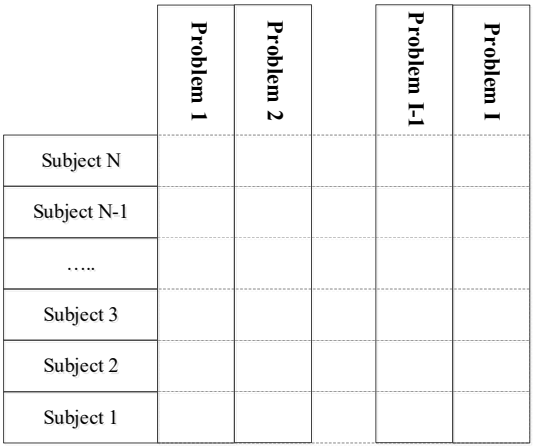

Figure 4.4. The process to design visual lecture according to problem.

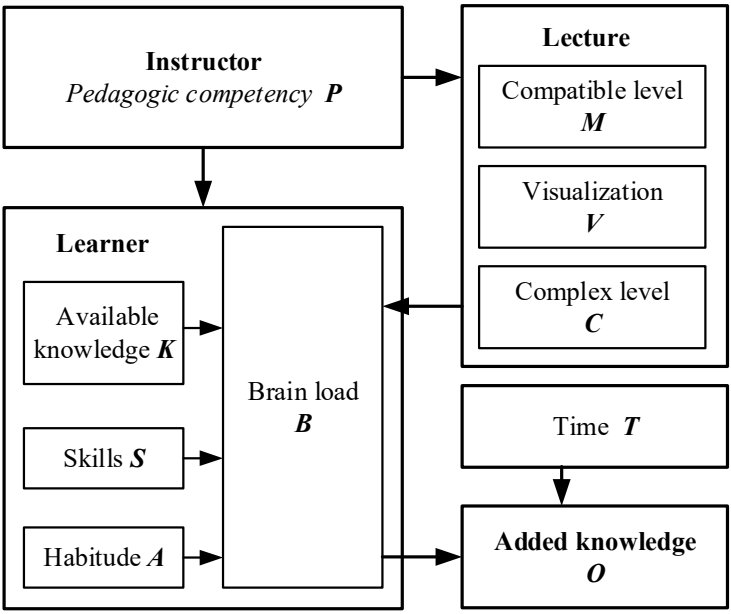

Figure 4.5. The model to evaluate the efficiency of the visually-guided-discovery system for a lecture.

The visually-guided-discovery-learning system is evaluated by the efficiency of each lecture with respect to each learners' group and instructor by comparing learners' achievement knowledge with mobilized brain load and learning time.

Conclusion of the chapter 4

The thesis proposed the model of visually-guided-discovery model combining with the characteristics of collaborative learning and problem learning. The model of visually-guided-discovery learning is in accordance with not only the development of society, data and information source, internet, but also necessary to train informatics team.

CONCLUSION

The scientific achievements of the Thesis. The thesis achieved the following results:

1. ***Dataset.*** *The thesis formed the dataset of the competency of examinees who have participated in the exam of national excellent students in informatics, including winners and nonwinners.*

The thesis set up the competency feature vector model of examinees, comprising learning features and non-learning features. The approach of reply matrix and reply 3D cube are proposed to quantify the sets of competency feature variables of competency space and the competency feature components of surveyed person. The result is 2 datasets: the dataset 2024 comprised 221 examinees (120 winners and 101 nonwinners) who participated in national informatics excellent student exam in school years 2020-2021, 2021-2022, 2022-2023; the dataset 2025 comprised 497 examinees (360 winners and 137 nonwinners) who participated in national informatics excellent student exam in school years 2021-2022, 2022-2023, 2023-2024.

The process setting up the dataset of the thesis is smart, suitable for the Winner-domain and Winner-cosin approaches and adaptive with examinees of national excellent student exams. The dataset formed by the thesis is only applied for the competitions experimented in the thesis. However, the approaches forming dataset applied in the thesis may be applied for other competitions by changing the number of instances, the number and contents of data variables according to the request of the competition.

2. ***The model of Winner-domain combining with Winner-cosin.*** *The thesis formed the model selecting excellent students for informatics team of a high school to participate in national excellent student exam in informatics by combining the proposed algorithms of Winner-domain and Winner-cosin.*

In the light of the philosophy “person near ink is darkened, person near light is brightened” and the strategy “Better to exclude than wrongly include”, the thesis formed the model combining the

algorithms of Winner-domain and Winner-cosine, proposed by the thesis, to select members of informatics team. The thesis was experimentally applied in school years 2023-2024 and 2024-2025.

3. ***The model of visually-guided-discovery learning.*** *The thesis proposed the model of visually-guided-discovery learning.*

The thesis combined discovery learning model with collaborative learning and project learning models, … and the theories of graph, human cognition load in the process acquiring and storing knowledge, along with visual technique to form the process and the method editing visual lecture for the model of visually-guided-discovery learning. The thesis also formed the mathematical model to evaluate the efficiency of each visual lecture.

The continued works of the thesis.

- In the reality, the thesis will update the values of competency variables of new competitions, and analyze competency variables for other disciplines, other competitions, as well as other machine learning models.
- For the science, the thesis will develop the method forming educational data; enhance the process of Good Winner-domain, Winner-cosin and decision tree; experimentally evaluate the efficiency of visually-guide discovery learning for a course.

PUBLISHED ARTICLES RELATING TO THE THESIS

[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 Exam," 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).