An Approach to Setting up Dataset for Winner-Domain Algorithm Selecting Teamers Participating in Vietnam National Competitions

Hong Thi Nguyen, Cam Ngoc Thi Huynh & Phuoc Vinh Tran

Thu Dau Mot University

ABSTRACT

The dataset of competitors' performance in some recent years is the input of winner-domain algorithm selecting good student as member of a subject team participating in Vietnam national prize. The data of learning performance features may be retrieved from the competitors' learning outcomes. Meanwhile, the collection of non-learning performance feature data of competitors meets challenges relating to the individual right and interviewees' cooperation for answering questionnaire. This research analyzed competitors' skills and lifestyle into very simple questions with very easy answer mode, without the deep relation to their individual lifestyle. The reply-matrix and reply-cube are formed to map simple answers of questionnaire reply into quantitative values. This approach may be flexibly applied to easily form the datasets of competitors' performance for various subjects in different years and at different levels of competitions.

Keywords: student performance, student competition, winner-domain, selecting teamers.

Classification: LCC Code: QA76.9.D343

Language: English

Great Britain Journals Press

LJP Copyright ID: 975812
Print ISSN: 2514-863X
Online ISSN: 2514-8648

London Journal of Research in Computer Science & Technology

Volume 25 | Issue 1 | Compilation 1.0

An Approach to Setting up Dataset for Winner-Domain Algorithm Selecting Teamers Participating in Vietnam National Competitions

Hong Thi Nguyen[α], Cam Ngoc Thi Huynh[σ] & Phuoc Vinh Tran[ρ]

ABSTRACT

The dataset of competitors' performance in some recent years is the input of winner-domain algorithm selecting good student as member of a subject team participating in Vietnam national prize. The data of learning performance features may be retrieved from the competitors' learning outcomes. Meanwhile, the collection of non-learning performance feature data of competitors meets challenges relating to the individual right and interviewees' cooperation for answering questionnaire. This research analyzed competitors' skills and lifestyle into very simple questions with very easy answer mode, without the deep relation to their individual lifestyle. The reply-matrix and reply-cube are formed to map simple answers of questionnaire reply into quantitative values. This approach may be flexibly applied to easily form the datasets of competitors' performance for various subjects in different years and at different levels of competitions.

Keywords: student performance, student competition, winner-domain, selecting teamers.

Author σ ρ: Institute o App ie Mec anics an Informatics, Ho Chi Minh City, Viet Nam.
α ρ: Thu Dau Mot University, Binh Duong, Viet Nam.
σ: Graduate University of Science and Technology, Ha Noi, Viet Nam.

I. INTRODUCTION

Beginning new school year, the leaders of high schools set up teams of subjects (subject teams), each of which is assigned to a teacher who acts as the team coach. The coach of a subject team sets up the team by choosing good students of the school as teamers (member of team) and deeply trains them before participating in national competitions.

Traditionally, the subject coach tests the good students of 11th and 12th grades to select the best students for his/her team. In some recent years, each teamer winning a national prize (winner) not only obtains the worth awards by the ministry, the province, the school but also is welcomed to enter big universities with high grant. Accordingly, the selection of teamers for subject teams becomes a big pressure towards subject coaches.

So far, the coaches of teams remain the traditional approach to find out teamers. Their feelings and students' learning outcomes indicate students for them to select. This selection is generating some negative problems. The winner-domain approach to objectively selecting teamers participating in National Prizes is developed according to the Vietnamese proverb "A man is known by the company he keeps" [1]. In other words, the winners of the same prize are similar in performance, including learning and non-learning performance.

The input of winner-domain algorithm is the dataset of competitors' performance features [1]. The learning performance features can be retrieved from the database of learning outcomes. Meanwhile, the non-learning performance data of competitors is a great challenge to coaches applying winner-domain algorithm for selecting teamers. The problem to reveal is how to adequately collect competitors' data of non-learning performance without offending their private life.

Moreover, the characteristics of some subject prize is different from another subject, from another time, and from another location. This research proposes a smart approach to setting up various datasets for selecting teamers participating in different prizes, without offending private life as well as not taking much time of interviewees. The idea to solve the challenges is to form the reply-matrix to transform the very simple answers of interviewees into numbers to set up smart dataset.

The article is structured as follows. The 2nd section reviews the performance features, especially non-learning performance including skills and lifestyle. The 3rd section proposes the approaches reply-matrix and reply-cube to collecting and processing competitors' data of non-leaning performance. The 4th section results in the dataset as a multivariable data table of competitors' non-learning performance. Finally, the conclusion in the 5th section summarizes the result of the research.

II. PERFORMANCE FEATURES

In the face of exam problem in the context of national excellent student competition, competitor is under intense pressure to solve challenging exam questions within limited time. In reality, some competitors of better learning performance may still fail due to factors beyond knowledge alone. It is considered that the competitors who win prizes not only have better learning performance but also possess good non-learning performance. In addition to academic knowledge, the skills, attitude and behavior are also the factors necessary for competitors' success [2-4].

For doing exam problems, competitors not only mobilize their own knowledge, but also automatically utilize skills necessary to cognize, thoroughly understand the challenges of the problem to defeat their obstacles as well as overcome unusual impacts from outside. The skills of passion, self-teaching, self-confidence, cognition, critical thinking, creative thinking, time management and the lifestyle affects competitors' behavior while doing exams. This research studies the non-learning factors impacting competitor's result.

2.1 Skills

The students' skills are not only native and heritable [5] but also affected by education, familial and social environment [1]. While doing exam problems, competitors face a number of pressure not only from the challenges of the problems, but also from the family, school, and social relations. Accordingly, they need possess necessary skills to overcome challenges of exam problems, potential facts, and their mentality.

Passion [6]. Winners of a competition are students who have had strong desire to take part in competition and win prize. Everyday, they can ignore other works to take all time for solving difficult problems of competition subject.

Self-Teaching [7, 8]. Winners of a competition are students proactive in learning to develop knowledge. Competitors are always passionate about academic issues of the subject, they self-motivate their learning, self-finding documents, self-discover new topics and approaches to solve new difficult problems.

Self- Confidence Skill [9]. The self-confidence skill enables competitors to deal with the challenge of exam problem. A competitor of high self-confidence may result in haughtiness and easily fail at competition. A competitor need own self-confidence skill without haughtiness.

Cognitive Skill [5, 10-15]. The cognitive skill enables competitors to acquire and process data from real world by the senses. The cognitive skill refers to the progress transforming information into knowledge to store in long-term memory, to the activities of storing and retrieving knowledge in the memory. Students of higher cognitive skills are more likely to achieve better results in learning. Competitors need possess good cognitive skill to well understand exam problems and recognize their challenges.

Critical Thinking Skill [15-20]. The critical thinking skill enables competitors to well understand a matter by reasoning, analyzing, evaluating, and commenting in the light of some perspective or theory relating to usual or unusual questions, situations and problems. A well critical thinker owns the skill receiving all various views on some matter to find out the answer most suitable for his decision. Competitors apply critical thinking skill to well understand exam problems and right recognize their challenges.

Creative Thinking Skill [16, 21-24]. The creative thinking skill enables competitors to generate new ideas for better making something or better processing an existent matter with a different and nontraditional approach. In most cases, the skills of creative thinking and critical thinking together with the activities of identifying, evaluating, analyzing and synthesizing, reasoning generate or develop new ideas to solve complex problems. In the face of hard exam problem, competitors concurrently mobilize critical and creative skills to generate the best solution.

Time Management Skill [25]. The time management skill refers to competitors' utilization of time while doing exam problems. The time management skill assists competitors in designing the reasonable plan to best solve the whole exam problem within the limited time frame.

2.2 Lifestyle

At high schools, competitors have been deeply trained at least one year before competition. Personal customs and family life strongly impact on the competitors' preparation for the mentality facing exam problems in competition context. Accordingly, the competitor's lifestyle is also considered as important factors affecting competition result.

Daily Time Usage [25]. Competitors are necessary to balance the amount of time spent on learning and relax with a reasonable timetable. Before competition, competitors have spent daily time in increasing academic knowledge, strengthening skills facing the possibilities of exam problems, and keeping reasonable relax time, concurrently. A competitor who has not the capacity for reasonable utilization of daily time is difficult to win prize.

Family Life [25]. The family life has strongly affected competitors' increase in knowledge, skills to prepare for participating in examination. The good learning tradition and kind interest of family morally encourage competitors in winning prize. The family of high income has the great capacity for assisting competitors in deeply training and participating in prizes.

Love. In reality, some event on competitor' love may affect his/her exam result. A competitor unlucky in love is very difficult to win prize at competition.

III. COLLECTION AND PROCESSING OF PERFORMANCE DATA

The competitor's performance is mathematically represented as a multivariable vector in multidimensional performance space, including learning feature variables and non-learning feature variables. The dataset applied for the winner-domain algorithm is a set of performance vectors of competitors belonging to the same subject of a prize. This article proposes an approach to form the non-learning performance dataset of competitors, including winners and non-winners of Vietnam national prizes on a subject for some recent years.

3.1 Questionnaires

Each year, the national excellent student competition in Vietnam is organized with various subjects such as information, mathematic, physic, chemistry, language, history, geography, and so on. The competitors are asked about knowledge and skills suitable for each yearly subject. The winner-domain approach may be applied for prizes of various subjects, each application for a subject of a prize in a year need a suitable dataset. The datasets applying for competitions of different subjects in different years at different levels are different.

In reality, it is impossible to collect the data of competitors' performance features by testing. With the respect of interviewees' personal life, this research has collected data by questionnaire with very simple questions and the request for very easy feedback. The non-learning performance features are analyzed into several simple questions not deeply referring to personal life of interviewees, each of which is studied as a performance factor and represented as a performance variable. The questionnaire is sent to some competitors of national prizes in two recent years.

The questions are flexibly edited based on editor's cognition about the relation between performance factors and life activities, the position of prize (province or nation) as the importance of the prize, the year and the subject of the competition. Each questionnaire can be edited with a number of suitable questions. Each question of questionnaire is composed of several responses with the symbols of a, b, c, d, e, f, g, h, i which interviewees can easily understand and fast answer by signing a simple tic symbol (X) to the convenient response.

3.2 Reply-Matrix

This research proposes the reply-matrix approach to processing the interviewees' replies of questionnaires. Reply-matrix is designed as a data table (Table 1), of which each line is in proportion to a question, each column is in proportion to responses of questions. This approach transforms each questionnaire reply into a reply-matrix as the table 1. Each questionnaire reply is mapped into a reply-matrix, where each response with the answer of X is assigned the number 1 to the corresponding position on the reply-matrix. The responses without answer are assigned the number 0 to the corresponding positions on the reply-matrix.

Table 1: Reply-matrix of the interviewee m | m=1,2,..,M

Questions	Responses of the interviewee m								
1	a_1^m	b_1^m	c_1^m	d_1^m	e_1^m	f_1^m	g_1^m	h_1^m	i_1^m
...	...	...	...	...	...	...	...	...	...
s	a_s^m	b_s^m	c_s^m	d_s^m	e_s^m	f_s^m	g_s^m	h_s^m	i_s^m
...	...	...	...	...	...	...	...	...	...
S	a_S^m	b_S^m	c_S^m	d_S^m	e_S^m	f_S^m	g_S^m	h_S^m	i_S^m

3.3 Reply-Cube

The model of reply-cube is formed to process all questionnaire replies. All reply-matrices of M interviewees are joined as a cube (Figure 1). Each line is in proportion to a performance variable $\mathrm{F}_s|\mathrm{s}=1,..,\mathrm{S}$ and each column is in proportion to a value $\mathrm{f}_{s.\beta} \in [0,10] \subset \mathbb{R} \}\beta=1,2,..\}$ of performance variables.

$$\mathrm{F}_s = \{\mathrm{f}_{s.\beta} \in [0,10] \subset \mathbb{R} \}\beta=1,2,..\}|\mathrm{s}=1,..,\mathrm{S} \tag{1}$$

The values $\{\mathrm{f}_{s.\beta}\}\beta=1,2,..\}$ of the performance variable $\mathrm{F}_s|\mathrm{s}=1,..,\mathrm{S}$ is calculated as follows.

$$\mathrm{f}_{s.1} = \frac{\delta}{M}\sum_{1}^{M} a_s^m \mid a_s^m \in \{0,1\} \mid m = 1,2,..,M; s = 1..,S \tag{2}$$

$$\mathrm{f}_{s.2} = \frac{\delta}{M}\sum_{1}^{M} b_s^m \mid b_s^m \in \{0,1\} \mid m = 1,2,..,M; s = 1..,S$$

$$\mathrm{f}_{s.3} = \frac{\delta}{M}\sum_{1}^{M} c_s^m \mid c_s^m \in \{0,1\} \mid m = 1,2,..,M; s = 1,..,S$$

$$f_{s.4} = \frac{\delta}{M}\sum_{1}^{M} d_s^m \mid d_s^m \in \{0,1\} \mid m = 1,2,..,M; s = 1,..,S$$

$$f_{s.5} = \frac{\delta}{M}\sum_{1}^{M} e_s^m \mid e_s^m \in \{0,1\} \mid m = 1,2,..,M; s = 1,..,S$$

$$f_{s.6} = \frac{\delta}{M}\sum_{1}^{M} f_s^m \mid f_s^m \in \{0,1\} \mid m = 1,2,..,M, s = 1,..,S$$

$$f_{s.7} = \frac{\delta}{M}\sum_{1}^{M} g_s^m \mid g_s^m \in \{0,1\} \mid m = 1,2,..,M; s = 1,..,S$$

$$f_{s.8} = \frac{\delta}{M}\sum_{1}^{M} h_s^m \mid h_s^m \in \{0,1\} \mid m = 1,2,..,M, s = 1,..,S$$

$$f_{s.9} = \frac{\delta}{M}\sum_{1}^{M} i_s^m \mid i_s^m \in \{0,1\} \mid m = 1,2,..,M, s = 1,..,S$$

The number δ is customized to standardize the values of $f_{s.\beta} \mid f_{s.\beta} \in [0,10] \subset R\}\beta=1,2,..|s=1,..,S$.

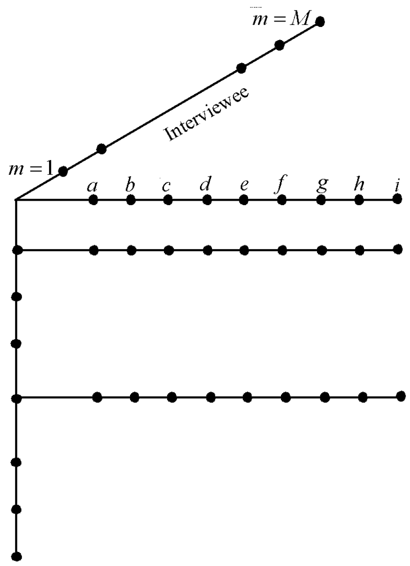

Fig. 1: Reply-cube to define the performance variables and determine the performance vectors of interviewees.

IV. SMART DATASET OF NON-LEARNING FEATURES

4.1 Performance Vector

The non-learning performance of an interviewee or a student $m \mid m = 1,..,M$ is represented as a vector $\vec{F}_m = (f_1^m, .., f_S^m) | m = 1,2,..,M$ which is determined as follows.

$$f_s^m = a_s^m.f_{s.1} + b_s^m.f_{s.2} + c_s^m.f_{s.3} + d_s^m.f_{s.4} + e_s^m.f_{s.5} + f_s^m.f_{s.6} + g_s^m.f_{s.7} + h_s^m.f_{s.8} + i_s^m.f_{s.9}$$

$$\text{for} \quad s = 1,..,S \mid m = 1,..,M \tag{3}$$

4.2 Smart Dataset

The dataset is formed as a data table of interviewees' performance vectors

$$\vec{F}^m = (f_1^m, .., f_S^m) | m = 1, 2, .., M \tag{4}$$

Table 2: The dataset of interviewees' performance variables

Interviewee	Performance variables s						
m = 1	f_1^1	...	...	f_s^1	...	...	f_S^1
...		...	...	...	...	...	...
m	f_1^m	...	...	f_s^m	...	...	f_S^m
...	...	...	...	...	...	...	...
m = M	f_1^M	...	...	f_s^M	...	...	f_S^M

V. CONCLUSION

The approaches of reply-matrix and reply-cube enable to set up the non-learning dataset of competitors for winner-domain algorithm selecting teamers participating in excellent student prize for each subject, each year and each level. These approaches collect the data on non-learning performance features of competitors by questionnaires with simple questions and responses easy to answer.

Winners of a prize own performance features suitable for the subject, the year and the level of the competition. The mapping of questions of questionnaire to performance variables customizes the dataset for applying winner-domain approach to various competitions. The reply-matrix and reply-cube are flexible approaches to smartly form the input dataset of winner-domain algorithm.

REFERENCES

1. C. N. T. Huynh, P. V. Tran, and T. V. Tran, "The Approach of Winner-domain to Selecting Members for High School Team Participating in National Excellent Student Competition," Mobile Networks and Applications, pp. 1-8, 2024.
2. H. T. Nguyen, A. V. T. Tran, T. A. T. Nguyen, L. T. Vo, and P. V. Tran, "Multivariate Cube for Representing Multivariable Data in Visual Analytics," in Context-Aware Systems and Applications, Thu Dau Mot, Viet Nam, 2016, vol. LNICST, pp. 91-100: Springer, 2016.
3. E. N. Khabibah, M. Masykuri, and Maridi, "The Effectiveness of Module Based on Discovery Learning to Increase Generic Science Skills," Journal of Education and Learning, vol. 11, no. 2, pp. 146-153, 2017.
4. H. T. Nguyen, A. V. T. Tran, T. A. T. Nguyen, L. T. Vo, and P. V. Tran, "Multivariate cube integrated retinal variable to visually represent multivariable data," EAI Endorsed Transactions on Context-aware Systems and Applications, vol. 4, no. 12, pp. 1-8, 2017.
5. M. Malanchini, K. Rimfeld, A. G. Allegrini, S. J. Ritchie, and R. Plomin, "Cognitive ability and education: how behavioural genetic research has advanced our knowledge and understanding of their association," Neuroscience and Biobehavioral Reviews, pp. 1-52, 2020.
6. A. L. Duckworth, GRIT: The power of passion and perseverance. New York: Scribner, 2016, p. 352.
7. A. S. Hashim, W. A. Awadh, and A. K. Hamoud, "Student Performance Prediction Model based on Supervised Machine Learning Algorithms," in 2nd International Scientific Conference of Al-Ayen University, 2020, vol. 928, p. 032019: IOP Publishing.
8. G. Mahajan and B. Saini, "Educational Data Mining: A state-of-the-art survey on tools and techniques used in EDM," International Journal of Computer Applications &

Information Technology, vol. 12, no. 1, pp. 310-316, 2020.
9. M. R. Ramdhani, "Discovery Learning with Scientific Approach on Geometry," Journal of Physics: Conference Series, vol. 895, 2017.
10. I. Damyanov and N. Tsankov, "The Role of Infographics for the Development of Skills for Cognitive Modeling in Education," International Journal of Emerging Technologies in Learning (iJET), vol. 13, no. 1, pp. 82-92, 2018.
11. J. Casagrand and K. Semsar, "Redesigning a course to help students achieve higher-order cognitive thinking skills: from goals and mechanics to student outcomes," Advances in physiology education, vol. 41, no. 2, pp. 194-202, 2017.
12. J. Bietenbeck, "Teaching Practices and Cognitive Skills," Labour Economics, vol. 30, no. 2014, pp. 143-153, 2014.
13. H. Tuaputty, M. Leasa, A. D. Corebima, and J. R. Batlolona, "The Correlation between Critical Thinking Skills and Cognitive Learning Outcomes," Ilkogretim Online, vol. 20, no. 1, pp. 302-317, 2021.
14. T. Bayne et al., "What is cognition?," Current Biology, vol. 29, no. 13, pp. R608-R615, 2019.
15. K. Shatri and K. Buza, "The Use of Visualization in Teaching and Learning Process for Developing Critical Thinking of Students," European Journal of Social Sciences: Education and Research, vol. 4, no. 1, pp. 71-74, 2017.
16. K. Yuliani and S. Saragih, "The Development of Learning Devices Based Guided Discovery Model to Improve Understanding Concept and Critical Thinking Mathematically Ability of Students at Islamic Junior High School of Medan " Journal of Education and Practice, vol. 6, no. 24, pp. 116-128, 2015.
17. E. R. Lai, "Critical Thinking: A Literature Review," Pearson's Research Reports, vol. 6, no. 40, pp. 1-41, 2011.
18. U. Toharudin, "Critical Thinking and Problem Solving Skills: How these Skills are needed in Educational Psychology?," International Journal of Science and Research vol. 66, no. 3, pp. 2004-2007, 2017.
19. G. R. Akramova, "Modern Approaches to Development Critical Thinking of Students," Eastern European Scientific Journal vol. 5, 2017.
20. J. Maynes, "Critical Thinking and Cognitive Bias," Informal Logic, vol. 35, no. 2, pp. 183-203, 2015.
21. O. W. A. Gafour and W. A. S. Gafour, "Creative Thinking skills – A Review article," Journal of Education and E-Learning, vol. 4, pp. 44-58, 2020.
22. M. H. Rahman, "Using Discovery Learning to Encourage Creative Thinking," International Journal of Social Sciences & Educational Studies, vol. 4, no. 2, pp. 98-103, 2017.
23. K. ÜLGER, "The Relationship between Creative Thinking and Critical Thinking Skills of Students," Hacettepe Universitesi Egitim Fakultesi Dergisi-Hacettepe University Journal of Education, vol. 31, 2016.
24. S. Suwandari, M. Ibrahim, and W. Widodo, "Application of Discovery Learning to Train the Creative Thinking Skills of Elementary School Student," International Journal of Innovative Science and Research Technology, vol. 4, no. 12, 2019.
25. A. Khan and S. K. Ghosh, "Student performance analysis and prediction in classroom learning: A review of educational data mining studies," Educational and Information Technologies, vol. 26, no. 1, pp. 205-240, 2021.