An Approach to Selecting Students Taking Provincial and National Excellent Student Exams

Cam Ngoc Thi Huynh[2,3], Hiep Van Nguyen[1(✉)], Phuoc Vinh Tran[1,2(✉)], Diu Ngoc Thi Ngo[1], Trung Vinh Tran[4(✉)], and Hong Thi Nguyen[1]

[1] Thudaumot University (TDMU), Thu Dau Mot, Binhduong, Vietnam
{hiepnv,phuoctv,diuntn,hongnt.ktcn}@tdmu.edu.vn,
Phuoc.gis@gmail.com
[2] Institute of Applied Mechanics and Informatics, Ho Chi Minh City, Vietnam
[3] Graduate University of Science and Technology, Ha Noi, Vietnam
[4] Fayetteville State University, Fayetteville, NC, USA
ttran1@uncfsu.edu

Abstract. The provincial and national excellent student exams reserved for Vietnam high schools are organized every year. These exams are the opportunities of high schools to acquire achievements. The problem to be solved by schools is how to select students for excellent student team in each discipline. This research considers that the students of the performance similar to winners in recent years are more likely to win prize. Mathematically, each student or winner is represented as a vector of performance of which features are influenced by the learning and non-learning factors. The vectors representing winners of earlier exams form winner-domain which is determined by its centroid, limiting distance, and limiting tendency. The students from classes are selected for the team of a discipline when their performance vectors are within the winner-domain of the discipline.

Keywords: Excellent student exam · Student selection · Student performance · Performance features · High school education

1 Introduction

In Vietnam, excellent student exams are organized to award prizes at provincial and national levels every year. Several provincial or gifted high schools founded gifted classes to deeply train selected students in each discipline. Every year, each school selects excellent students to form discipline teams for provincial and national exams. The selection is being created based on students' learning outcomes and teachers' perceptibility. The problem to reveal is how to objectively select students for teams.

The idea is that the students of performance features similar to prize-winners in a discipline are most likely to win prize of the discipline. This research approaches the theory of machine learning to selecting students of competency and academic tendency

C. V. Phan and T. D. Nguyen (Eds.): ICCASA 2022, LNICST 475, pp. 156–161, 2023.
https://doi.org/10.1007/978-3-031-28816-6_12

convenient for excellent student exams. Each student or winner is mathematically represented as a vector of performance. For a discipline of a high school, the students of performance similar to winners' performance are selected for excellent student team.

The paper is structured as follows. The next section refers to the context of the research and the related works. The third section divides the features influencing student performance into learning and non-learning factors and mathematically represent the performance as a vector of features. The fourth section applies the method of supervised machine learning to form winner-domain and select students for teams taking excellent student exams. Finally, the fifth section summarizes the contents of the research as well as the difficulty in fact.

2 Context and Related Works

2.1 Context

The excellent student exams are outstanding events of Vietnamese Education sector, every year. The national and provincial exams are not only interested by gifted high schools, but also by other high schools. The schools' leaders expect to congratulate prize-winners from the teams of their schools because these wins significantly contribute to the achievements of the school. This view results in the strong investment of schools in selecting and training students for teams. The students of good performance in the discipline are more likely to win prizes at the exams.

Just at the beginning of each academic year, high schools select excellent students from 10^{th}, 11^{th}, and 12^{th} grades of the disciplines as Information Technology, English language, Geography, etc. to form teams in each discipline. The teachers of each discipline select students for their teams by feeling based on learning outcomes, skills and attitude in learning of the students. The students of teams are taught with specific curricula to take the exams at the end of academic year. Consequently, the selection of students for teams plays an important role in the win at exams.

2.2 Related Works

In recent year, the educational data exploitation is a topic attracting several authors to research for various purposes, specially for improving teaching quality, organizing classroom, intervening student performance early [1–4]. The works focus on the factors influencing the performance of students. Several authors have tried to identify and find out the factors impacting on student performance [2, 3]. They analyze the factors influencing student performance in classroom learning before and after course commencement [2].

Some authors consider that the features of knowledge, individual skills and attitude in learning are the factors influencing student performance [5, 6]. The student's knowledge concerns the teaching of previous school, neighborhood, and age [2]. The student's skills and attitude as self-motivation [7], self-confidence [8], self-reliance [8], self-finding [9], self-investigation [9], self-analysis [10], critical thinking [11], creative thinking [12], interaction [12] also influence student performance. In addition, the socio-economic and demographic background, the family, and behavior of students also impact on their performance [2, 13].

The data mining and machine learning techniques have been significantly contributing to the discovery of values hide in educational data. Some authors have applied the approaches of classification, clustering, regression, decision trees, neural networks, nearest neighboring to analyze learning, performance, and the correlation between performance and impacting factors [14]. Others have applied the algorithms of supervised, unsupervised, semi-supervised, reinforcement leaning to discover student models, predict student performance [15–17].

3 Modelling Performance

The features influencing the performance are composed of the learning factors and non-learning factors. The learning factors can be estimated by learning history, while the non-learning factors composed of personality, family background, and socio-economic environment can be surveyed by questionnaires and/or the other information sources from family and society. Mathematically, the performance is represented as the vector of features, where the vector representing the performance of a student is called student-vector, symbolled as follows.

$$\mathrm{x}_n = [s_{1.n}, .., s_{i.n}, .., s_{I.n}]^{\mathrm{T}} = [s_{1.n}; ..; s_{2.i}; ..; s_{I.n}]$$

where:

$\mathrm{x}_n | n = 1, 2, \ldots$: the student-vector of the student n.
$s_{i.n} | i = 1, 2, \ldots, I$: the feature i of the student n.

The winners of national or provincial excellent student exams in the same discipline are similar in performance [2]. This idea infers that the students of performance similar to winners are more likely to win prize. In this research, the students of the performance similar to the performance of winners in a discipline are selected to be continuously trained for taking the next exams in the discipline. Mathematically, all winner-vectors representing the performance of the winners in the same discipline form the performance domain of the winners in the discipline, called winner-domain; the students who have their performance vectors within winner-domain are selected for team.

4 Selecting Students for Taking Excellent Student Exams

4.1 The Mathematical Features of Winner-Domain

The winner-domain of a discipline is determined by three mathematical features, including the domain centroid, the limiting distance, and the limiting tendency. The domain centroid is the centroid vector of the winner-domain, the limiting distance is the maximal distance from the domain centroid to all vectors of the domain, the limiting tendency is the maximal angle formed by the domain centroid with all vectors of the domain. The following process determines the mathematical features of a winner-domain.

Input: the winner-domain of a discipline is represented as the set W of the winner-vectors of the discipline.

Output: the centroid vector w_c, the limiting distance $d_{\lim}$, and the limiting tendency $\theta_{\lim}$ of the winner-domain W.

- *Step 1:* Define $W = \{w_1, .., w_n, .., w_N\}$ is the winner-domain, where each winner-vector is represented as $w_n = [s_{1.n}, .., s_{i.n}, .., s_{I.n}]^T = [s_{1.n}; ..; s_{i.n}; ..; s_{I.n}]$ which is the vector of the variables of features $s_{i.n}|i = 1, .., I; n = 1, .., N$
- *Step 2:* Determine the centroid w_c of the vector set W.

  The centroid $w_c = [s_{1.c}; ..; s_{i.c}; ..; s_{I.c}]$ of the vector set W is determined by:

$$s_{i.c} = \frac{\sum_{n=1}^{N} s_{i.n}}{N} | i = 1, 2, .., I$$

- *Step 3:* Determine the limiting distance $d_{\lim} = (d_n)_{\max} = (d(w_c, w_n))_{\max}$ of the vector set W.

  The distance $d_n = d(w_c, w_n)$ from the centroid w_c to the vector $w_n|n = 1, 2, .., N$ is defined by Euclidean distance as follows.

$$d_n = d(w_c, w_n) = \sqrt{(s_{1.c} - s_{1.n})^2 + .. + (s_{i.c} - s_{i.n})^2 + .. + (s_{I.c} - s_{I.n})^2}$$

- *Step 4:* Determine the limiting tendency-angle $\cos\theta_{\lim} = (\cos\theta_n)_{\min} = (\cos(w_c, w_n))_{\min}$ of the vectors in the vector set W.

  The tendency-angle θ_n of a vector w_n, which is the angle formed by the centroid w_c and the vector $w_n|n = 1, 2, .., N$, is determined by:

$$\cos\theta_n = \cos(w_c, w_n) = \frac{w_c.w_n}{|w_c||w_n|} = \frac{s_{1.c}s_{1.n} + .. + s_{i.c}s_{i.n} + .. + s_{I.c}s_{I.n}}{\sqrt{(s_{1.c})^2 + .. + (s_{I.c})^2}\sqrt{(s_{1.n})^2 + .. + (s_{I.n})^2}}$$

4.2 The Selection of Students for Excellent Student Team

All students in a discipline of school may be the candidates for the excellent student team in the discipline. Each candidate is represented as a performance vector of features. The following algorithm indicates the candidates of the performance similar to winners to be selected for the gifted class or the team taking national and provincial excellent student exams.

Input:

- The winner-domain features w_c, $d_{\lim}$, and $\theta_{\lim}$;
- The vector set $X = \{x_1, .., x_j, .., x_J\}$ represents candidates of whom the performance features are collected early.

Output: The vector set $Y = \{y_1, .., y_k, .., y_K\}$ represents the excellent candidates selected from the set X.

- *Step 5:* Calculate the distances d_j from x_j to the centroid w_c of the winner-domain for $j = 1, .., J$.

$$d_j = d(w_c, x_j) = \sqrt{(s_{1.c} - s_{1.j})^2 + .. + (s_{i.c} - s_{i.j})^2 + .. + (s_{I.c} - s_{I.j})^2}$$

- *Step 6:* Calculate the tendency-angle θ_j between the vector x_j and the centroid w_c of the winner-domain for $j = 1, .., J$.

$$\cos\theta_j = \cos(w_c, x_j) = \frac{s_{1.c}s_{1.j} + .. + s_{i.c}s_{i.j} + .. + s_{I.c}s_{I.j}}{\sqrt{(s_{1.c})^2 + .. + (s_{I.c})^2}\sqrt{(s_{1.j})^2 + .. + (s_{I.j})^2}}$$

- *Step 7:* Compare d_j with $d_{\lim}$ and θ_j with $\theta_{\lim}$.

  – If $(d_j - \delta d_{\lim} \leq 0) \wedge (\cos\theta_{\lim} - \delta\cos\theta_j \leq 0)$, then $x_j \mapsto y_k \in Y | j = 1, 2, .., J; k = 1, 2, .., K$
  – If $(d_j - \delta d_{\lim} > 0) \vee (\cos\theta_{\lim} - \delta\cos\theta_j > 0)$, then reject this candidate.

  where δ is an arbitrary number estimated by the number of students of the gifted class or the team.

5 Conclusion

This research approached supervised machine learning techniques to selecting excellent students of a high school for teams taking excellent student exams. The approach represented each student as a performance vector of features. The performance vectors of winners in a discipline are structured as the performance domain of winners defined by the domain-centroid, the limiting distance, and the limiting tendency. The student selected for the team has the performance vector within winner-domain, i.e. the distance from the student-vector to the domain-centroid is less than the limiting distance and the tendency of the student-vector is less than the limiting tendency.

The article presents the process selecting excellent students for the teams taking provincial and national excellent student exams in each discipline. The approach is being experimentally applied at a Gifted High School in Kiengiang province - Vietnam to form the team in the discipline of information technology for the next provincial and national excellent student exams. In fact, the data collection of non-learning factors of winners and students is difficult to completely carry out in the first instance. The methods for collecting the data of non-learning factors will be continuously discussed.

References

1. Romero, C.O., Ventura, S.A.: Educational data mining: a review of the state of the art. IEEE Trans. Syst. Man Cybern. Part C Appl. Rev. **40**(6), 601–618 (2010)
2. Khan, A., Ghosh, S.K.: Student performance analysis and prediction in classroom learning: a review of educational data mining studies. Educ. Inf. Technol. **26**(1), 205–240 (2020). https://doi.org/10.1007/s10639-020-10230-3
3. Alturki, S., Alturki, N.: Using educational data mining to predict students' academic performance for applying early interventions. J. Inf. Technol. Educ.: Innov. Pract. **20**(2021), 121–137 (2021)
4. Şahin, M., Yurdugül, H.: Educational data mining and learning analytics: past, present and future. Bartın Univ. J. Fac. Educ. **9**(1), 121–131 (2020)

5. Nguyen, H.T., Tran, A.V.T., Nguyen, T.A.T., Vo, L.T., Tran, P.V.: Multivariate cube for representing multivariable data in visual analytics. In: Cong Vinh, P., Tuan Anh, L., Loan, N., Vongdoiwang Siricharoen, W. (eds.) Context-Aware Systems and Applications. LNICST, vol. 193, pp. 91–100. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-56357-2_10
6. Nguyen, H.T., Tran, A.V.T., Nguyen, T.A.T., Vo, L.T., Tran, P.V.: Multivariate cube integrated retinal variable to visually represent multivariable data. EAI Endors. Trans. Context-Aware Syst. Appl. **4**(12), 1–8 (2017)
7. Druckman, D., Ebner, N.: Discovery learning in management education: design and case analysis. J. Manage. Educ. 1–28 (2017)
8. Ramdhani, M.R.: Discovery learning with scientific approach on geometry. J. Phys. Conf. Ser. **895** (2017)
9. Suwandari, S., Ibrahim, M., Widodo, W.: Application of discovery learning to train the creative thinking skills of elementary school student. Int. J. Innov. Sci. Res. Technol. **4**(12), 410–417 (2019)
10. Nguyen, H.T.: A model representing visually multivariable spatio-temporal data. Doctor of Philosophy, The faculty of Computer Science, University of Information Technology - Vietnam National University in Hochiminh City (2020)
11. Yuliani, K., Saragih, S.: 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. J. Educ. Pract. **6**(24), 116–128 (2015)
12. Simamora, R.E., Saragih, S., Hasratuddin: Improving students' mathematical problem solving ability and self-efficacy through guided discovery learning in local culture context. Int. Electron. J. Math. Educ. **14**(1), 61–72 (2019)
13. Zhang, X., Sun, G., Pan, Y., Sun, H., He, Y., Tan, J.: Students performance modeling based on behavior pattern. J. Ambient. Intell. Humaniz. Comput. **9**(5), 1659–1670 (2018). https://doi.org/10.1007/s12652-018-0864-6
14. Patil, J.M., Gupta, D.S.R.: Analytical review on various aspects of educational data mining and learning analytics. In: 2019 International Conference on Innovative Trends and Advances in Engineering and Technology (ICITAET), pp. 170–177. IEEE (2019)
15. Hashim, A.S., Awadh, W.A., Hamoud, A.K.: Student performance prediction model based on supervised machine learning algorithms. In: 2nd International Scientific Conference of Al-Ayen University, vol. 928, p. 032019. IOP Publishing (2020)
16. Li, N., Matsuda, N., Cohen, W.W., Koedinger, K.R.: A machine learning approach for automatic student model discovery. In: The 4th International Conference on Educational Data Mining, Eindhoven (2011)
17. Ayodele, T.O.: Types of machine learning algorithms. New Adv. Mach. Learn. **3**, 19–48 (2010)