IIAI Open Conference Publication Series
IIAI Letters on Institutional Research
Vol. 003, LIR159
DOI: https://doi.org/10.52731/lir.v003.159

Typology of Students Graduating from College Using Dropout Probabilities

Naruhiko Shiratori *

Abstract

This study uses dropout probability to identify the degree to which students who graduated came close to dropping out and to examine how students who graduated went from being close to dropping out to graduating. Dropout probability is a number presented by Shiratori that represents how likely a student is to drop out [3]. Although the dropout probability has been used to analyze the patterns of students who dropped out of school, the process by which students graduated has not been analyzed. In this study, the dropout probability was used to quantify the student status of students who graduated, semester by semester and student by student, and the student status up to graduation was summarized as a transition vector. The transition vectors were then typified using k-means to create patterns of graduated student states. Among the students who graduated, we found that about 14% were in a state in which they could have dropped out. By analyzing the above patterns, we were able to determine how many students were in poor condition and at what stage they were likely to graduate.

Keywords: Typology, Pattern, Dropout Probability, Clustering

1 Introduction

The overall university dropout rate in Japan is about 8%, according to a Yomiuri Shimbun survey [1]. The dropout rate at national universities is 3%, public universities 4%, and private universities 9%, with private universities faring worse. Overall, the dropout rate tends to be low, ranging from 3 to 9%, but it often exceeds 20% at some universities. According to a survey by the Ministry of Education, Culture, Sports, Science and Technology (MEXT), 57875 students drop out in one year, which means that dropping out of college is a waste of time and wages for students and a management problem for universities [2].

According to Seidman and colleagues, measures to prevent dropouts require early detection and early, intensive, and continuous intervention. Early detection of students at risk of dropping out requires understanding the dropout factors for each university and each student, and calculating the student's risk based on the student's behavior and data about the student. After that, depending on the calculated risk and factors, appropriate dropout prevention intervention measures must be implemented in an intensive and continuous manner.

In order to implement measures to the right students at the right time, it is necessary to understand the state transitions of current students, such as how they drop out and graduate. In this study, we will understand student state transitions from the time students enter college

* The Faculty of Management and Economics, Kaetsu University, Tokyo, Japan

until they graduate, focusing on graduates. In order to understand student status, the dropout probability proposed in Shiratori 2020 [3] is used to calculate the student status for each semester and summarize it as a transition vector until graduation. Next, by typifying the transition vectors, we will identify the state in which graduates ended up to graduation.

2 Related Works

Research that predicts in advance which students are at high risk and to what degree is dropout prediction research, which uses prediction formulas to determine student status. Kondo et al. propose a study state Bayesian network to estimate the enrollment status of students from their first year, second year, and third year, using variables known before enrollment, such as gender, faculty affiliation, and submission of pre-enrollment study assignments, and variables known after enrollment, such as semester attendance rate and GPA [5]. Since a Bayesian network is used, it is possible to have multiple objective variables and to estimate the state of enrollment at each stage from the first year to the third year. Shiratori et al. use a logistic regression model to define each student's likelihood of dropping out as a dropout probability per semester [3]. Instead of using a binary value to determine whether a student is likely to drop out or not, they are able to determine the degree to which a student is likely to drop out and calculate the pattern to drop out. Elsewhere, Shiratori et al. use a random forest regression model to typify student status and transitions within the spring semester of the first year using intra-semester attendance data [6]. As described above, the predictive equations are used to estimate student status at the university.

Previous dropout prediction studies tended to focus on students who dropped out and did not analyze the extent to which students who graduated came close to dropping out. The aforementioned study by Shiratori et al. also derived patterns for students who dropped out, but did not clarify how students who were likely to drop out but did not drop out continued until graduation. In this study, we will use dropout probability to understand the status of students, but unlike the previous study, we will focus on students who graduated and categorize them to clarify how many of them were likely to drop out and how they continued their studies until graduation.

3 Methods

A. Data and Varibales

The data used in this study covers students enrolled from 2012 to 2014 at University A in Tokyo (a single college of social sciences), and does not cover transfer students, re-enrollment students, early graduates, or graduate students. The data will be for 849 students, of which 639 students graduated. The objective and explanatory variables are as follows: two types of explanatory variables are used: one that can be obtained by the time of admission and one that can be obtained after admission.1. Objective variable: Whether the student dropped out after the semester in question

2. Explanatory variables 1 (obtainable by the time of admission): gender, international students, number of days absent from high school, grade point average in high school, full-time high school, correspondence high school

3. Explanatory variables 2 (obtained after admission): total credits up to the semester in question, number of credits obtained per semester, GPA per semester

B. Calculation of Dropout Probability

The dropout probability is a numerical expression of whether a student is likely to drop out after the relevant semester, as proposed by Shiratori 2020. We will calculate the dropout probability for each semester and for each student using the two types of variables described in the previous section: variables that are known before enrollment and variables that are known after enrollment. The formula to calculate the dropout probability for semester s for student m is as follows. The equation to calculate the dropout probability is a logistic regression model equation, where the first term is a constant term, the second term represents the effect of variables that can be obtained before enrollment, and the third term represents the effect of variables that can be obtained after enrollment.

$$\ln\left(\frac{p_s[m]}{1-p_s[m]}\right) = \alpha_{s,0} + \sum_{i=1}^{6} \alpha_{s,i} x_i[m] + \sum_{k=1}^{s} \sum_{j=1}^{3} \beta_{s,k,j} x_{k,j}[m]$$

Using the above equations and data, maximum likelihood estimation is performed for each semester, and coefficients α and β are calculated for each semester from the beginning of the spring semester of the first year (1-1) to the beginning of the fall semester of the fourth year (4-2), creating a model for each semester. Subsequently, by substituting student m's data for each semester, the dropout probability $p_S[m]$ of student m for that semester s is calculated.

ID	1-1	1-2	2-1	2-2	3-1	3-2	4-1	4-2
1	0.14	0.04	0.04	0.03	0.01	0.00	0.00	0.01
2	0.15	0.04	0.04	0.02	0.00	0.00	0.00	0.00
4	0.23	0.02	0.04	0.04	0.01	0.00	0.00	0.00
7	0.09	0.07	0.10	0.10	0.02	0.02	0.02	0.03
8	0.18	0.05	0.05	0.05	0.01	0.00	0.02	0.01
13	0.12	0.03	0.02	0.01	0.00	0.00	0.00	0.00
14	0.27	0.26	0.18	0.05	0.00	0.00	0.01	0.01
15	0.17	0.25	0.10	0.08	0.15	0.19	0.03	0.00
16	0.21	0.08	0.05	0.04	0.01	0.00	0.00	0.00
18	0.17	0.04	0.04	0.03	0.01	0.01	0.00	0.00
19	0.10	0.04	0.03	0.03	0.02	0.01	0.01	0.00
21	0.60	0.45	0.36	0.14	0.01	0.00	0.00	0.00
22	0.07	0.11	0.05	0.03	0.04	0.01	0.00	0.00
24	0.13	0.05	0.02	0.01	0.00	0.00	0.00	0.00
25	0.07	0.06	0.02	0.01	0.00	0.00	0.00	0.00

Figure 1: Example of Dropout Probability

Figure 1 shows the calculated dropout probability for each student. In Figure 1, the horizontal axis represents the semester and the vertical axis represents the student here, with high dropout probability in red and low dropout probability in blue. For example, the dropout probability at the beginning of the spring semester of the first year for student ID1 is 0.14, which is slightly high, but at the beginning of the fall semester of the first year it is 0.04, which is low, and at the beginning of the spring semester of the third year it is 0.01, which is low. As in this example, each student would have a semester dropout probability of 8.

C. Creation and Categorization of Transition Vectors for Each Student

Using the semester-by-semester dropout probabilities described above, we create a vector pm that represents the dropout probability transition for each student. pm has dropout probabilities from p_{1-1}[m] at the beginning of the spring semester of the first year to p_{4-2}[m] at the beginning of the fall semester of the fourth year, and each vector has 8 elements. The following will be the transition vector for student m.

$$p[m] = (p_{1-1}[m], p_{1-2}[m], ..., p_{4-2}[m])$$

Next, we use transition vectors to create a typology. The data used were transition vectors for each individual, and the typification method was k-means (k=5) The number of k was arbitrarily determined.

4 Experimental Results and Discussion

A. Difference in Dropout Probability between Graduates and Dropouts

Table 1 shows the mean and standard deviation by semester for students who graduated and those who dropped out. It can be seen that as the semester progresses, the mean and standard deviation of the dropout probability for students who graduated decrease and the standard deviation also decreases. For students who dropped out, the dropout probability increases from the spring semester of the first year and stabilizes above 0.6 from the spring semester of the second year.

Table 1: Mean and Standard Deviation of Dropout Probability

	Normal		Dropout	
Semester	Mean	SD	Mean	SD
1_1	0.28	0.09	0.33	0.12
1_2	0.19	0.17	0.53	0.30
2_1	0.15	0.17	0.61	0.31
2_2	0.12	0.16	0.63	0.33
3_1	0.09	0.14	0.63	0.34
3_2	0.06	0.13	0.61	0.31
4_1	0.04	0.12	0.63	0.30
4_2	0.03	0.10	0.65	0.29

In the spring semester of the first year, the difference between students who graduated and those who dropped out is about 0.05, but after that, the mean probability of dropping out is significantly higher for students who dropped out (Welch's t-test p<.01). However, when the mean-1 standard deviation of the dropout probability of students who dropped out is calculated, it is 0.23 for the fall semester of the first year, 0.31 for the spring semester of the second year, and 0.30 for the fall semester of the second year, indicating that many students have a dropout probability of about 0.2 to 0.3, even among students who dropped out. Furthermore, when the mean + 1 standard deviation of the dropout probability of graduating students is calculated, the mean + 1 standard deviation is 0.36 for the fall semester of the first year, 0.32 for the spring semester of the second year, and 0.27 for the fall semester of the second year, indicating that there are students who have a dropout probability of about 0.2 to 0.3 even if they have graduated. Thus, it can be seen that even students who dropped out have a dropout probability of about 0.3, even if they graduated.

B. Results of the typology

Figure 2 shows the average dropout probability per class. The horizontal axis represents the semester and the vertical axis represents the mean dropout probability per class. The number of students in each class is shown in Table 2, and the classes are named from Class 1 to Class 5 in descending order of number of students.

The dropout probabilities for all groups from class 1 to class 5 begin to decrease from the fall of the second year, and by the spring of the fourth year, all classes have low values (dropout probability is less than 0.1).

The class in which the dropout probability remains low from the spring of the first year to graduation is 1, with 364 out of 639 students (60.0%). Approximately 60% of the graduates in this group graduate without a dropout probability higher than 0.1. Class 2 is a group in which the dropout probability in the first year is 0.1 higher than Class 1 at the beginning of the spring of the first year but gradually decreases as the school year progresses, as in Class 1, with 185 students (29.0%) in this class. The total number of students in Class 1 and Class 2 is 549 (86.0%), which

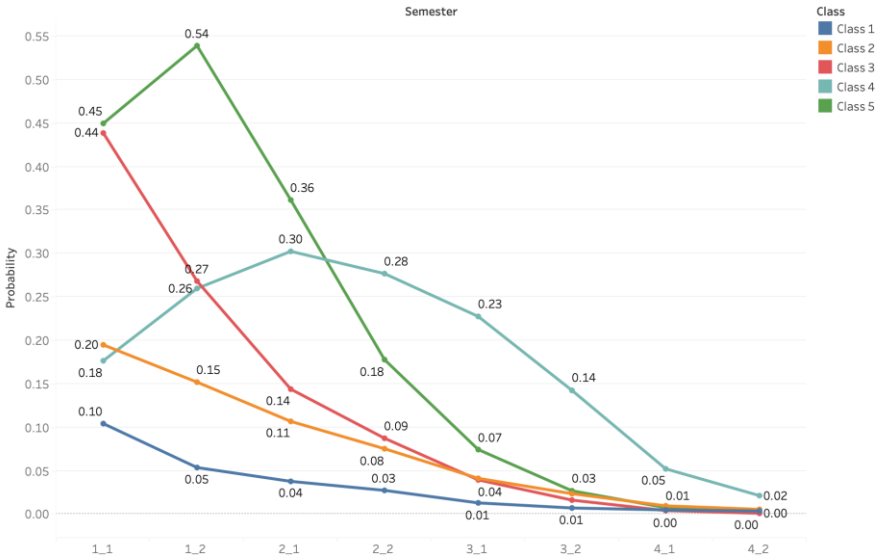

Figure 2: Dropout Probability Number of first-year spring semester credits of dropouts and graduates

means that 80% of the graduates are students who are doing well from the time they enter the school to the time they graduate.

On the other hand, students in Classes 3 through 5 do not progress as smoothly to graduation as students in Classes 1 and 2. Class 3, with 44 students (6.9%), has a high dropout probability of 0.44 in the spring semester of the first year and a relatively high dropout probability from there to the second year. Students in Class 3 are a group whose dropout probability was high in the first year, but decreased each semester as in Classes 1 and 2.

On the other hand, Class 4 and 5 are both groups in which the dropout probability temporarily increases in the spring semester of the first year and then decreases from a certain semester. Class 4, with 27 students (4.2%), is a group whose dropout probability increased from the first year, peaked in the spring semester of the second year, and improved after the fall semester of the second year. Class 5, with 19 students (2.9%), is a group in which the dropout probability is high in the first year and improves after the second year

5 Conclusion

In this study, we used dropout probability to understand the transition of student status from entering college to graduation, and typified the transition of student status with graduates. We identified the number of students who had graduated but were close to dropping out, and discussed how this was linked to graduation.

In the future, we would like to clarify the reasons why students who were at high risk were able to graduate. Since both classes 3, 4, and 5 were high-risk groups of students who were able to graduate, we would like to clarify why the risk was lowered. In addition, since the class classification method was fixed, it is necessary to seek the validity of the number of classes in the classification. Once the accuracy of the categorization is improved and the factors that reduce risk are understood, we believe that it will be possible to link this to effective dropout prevention measures.

Acknowledgement

This work was supported by JSPS KAKENHI Grant-in-Aid for Scientific Research C 20K02618.

References

[1] Yomiuri Shimbun Kyoiku Network, [University Competencies 2019] ”Daigaku no Jitsuryoku 2019 (in Japanese)”,Chuo Koron Shinsh, 2018.

[2] Ministry of Education, Culture, Sports, Science and Technology. [Survey on the status of student enrollment (drop-outs and leaves of absence) (as of the end of the 2021 academic year)] “Gakusei no shugakujokyo (Chutaisha, Kyugakusha) ni kansuru Chosa (Reiwa 3nendo matsu jiten) (in Japanese)”. 2022.

[3] Naruhiko Shiratori et al., [Making Dropout Patterns Using Transition of Dropout Probability]” Chutaikakuritsu no seni wo motiita chutaigakusei no ruikeika (in Japanese)”. Nihon

Kyoikukougakkai Ronbunshi (Japan journal of educational technology), vol 44, no.1, pp. 11-22, 2020.

[4] Seidman, Alan. "Taking Action: A Retention Formula and Model for Student Success." College Student Retention: Formula for Student Success. pp. 267–84. 2012.

[5] Nobuhiko Kondo and Toshiharu Hatanaka., [Modeling of Learning Process based on Bayesian Network] "Bayesian Network niyoru Shugakujoutai Suitei Model no Koutiku (in Japanese)". Nihon Kyoikukougakkai Ronbunshi (Japan journal of educational technology), vol. 41, no. 3, 2018, pp. 271–81.

[6] Naruhiko Shiratori, et al., [Typification of first-year spring semester study status using predicted GPA trends] "Yosoku GPA no suii wo mochiita 1nenji harugakki gakushu jotai no ruikeika (in Japanese)". Kyoiku System Johogakkaishi (Transactions of Japansese Society for Information and Systems in Education), vol. 39, no. 4, 2022, pp. 440–51.