The Exploration of Online Engagement Data in LMS as Predictors to E-Learning Outcomes

Ching-Rong LEE
National Open University, Taiwan
*crlee@mail.nou.edu.tw

Abstract: In this study, it is proposed an approach to utilize the students' online engagement data, in terms of the "counts", collected by the LMS. Data about 364 students who learned online throughout a semester was analyzed. Due to the skewed and peaked distribution, the negative binomial regression was applied to the data analysis. The test scores and time spent in e-learning produce the significant effects on the log of the counts of the LMS login, the counts of course studying, as well as the counts of the e-pages read. It was shown that using the count outcome variables can form the relationships with the predictors in a linear model.

Keywords: E-learning, Online engagement, Learning Management System

Introduction

American National Survey of Student Engagement (ANSS) addresses the concern of the amount of time and effort that students devoted to their studies and other educationally purposeful activities. The key concept of "Engaged Learning" can also be applied to e-learning (Thurmond, Wambach, and Connors, 2002). The disengaged e-learners is a challenge for the instructors who try to make extensive use of various pedagogies (Hiltz, 2004). If the e-courses had numerous students, the measures of the student online engaged behaviors became time consuming. Prior studies in e-learning adopt the survey-oriented approaches (Robinson & Hullinger, 2010), such as exploring students' attitudes about e-learning, overall students' satisfaction, and the participative effects on learning outcomes. It is argued that the most reliable evaluation of educational program's effectiveness is derived from performance-based measures as a data-driven approach (Kirkpatrick, 1994). This argument is well supported by the fact that learners' sense of engagement with courses is more dependent on their connection with learning materials than with the instructor or peers (Conrad, 2002). In this study, the way of students' online learning behaviors counted and tracked by the LMS (Learning Management System) is applied to be a data-driven method for e-learning educators.

Purpose and Research Questions

LMS is capable of meeting three pedagogical features: (a) a repository of course materials, (2) communication facilities, and (3) a platform for communication over the Internet. One advantage is often ignored by the instructors is the functions of data collection, in terms of counting, tracking, and recording of students' online behaviors. Most of the LMS counts the data, such as the occurrence of the login, the occurrence of the studying course materials, the occurrence of the pages read, the time spent in e-learning, and so on. When the dependent variable is a count variable, the Poisson or negative binomial distributions are commonly used to represent its distribution. It is in the hope that significant predictors could be identified by using this new analytic methods. Three research questions were formed: (1) which engagement data can be useful for prediction; (2) what the differences existed between the predictors' impacts, (3) which a regression model has a good fit of the data.

Research Methods: A Case Study and Regression Model

A course entitled "Introduction to Information Management" was offered by Taiwan National Open University Taiwan (TNOU). In spite of attending face-to-face group tutoring of total 8 hours on four weekends, e-learners studied via an e-learning platform through 18 weeks. Both of the midterm and the final exams in the format of the paper-pencil test were administered in proctor-based classrooms nationwide.

The Descriptive Data about the E-learners' Online Behaviors

There were 443 distance students nationwide enrolled this course. 54% were males. In the end of the semester, 64 students dropped, accounting for 14% of the drop rate. 15 students' records were excluded as outliers. Table 1 shows a great variability of the students' data. On the average, a student accessed this course for study 17 times, 83 pages, and 13 hours throughout 18 weeks. The counts of LMS login accumulated the prior and the login of other e-courses. It is noted that the positive skew and peaked kurtosis failed to represent a normal distribution (see Table 1 & Figure 1).

Table 1: The descriptive data of the students' backgrounds and online behaviors

N=364	Age	Level of Years	Midterm Scores	Final Scores	Count of LMS Login	Count of Freq. in Studying	Count of Pages Read	Time(hr) Spent E-learning
Minimum	18.0	1.0	8.0	2.0	0	0	0	0
Maximum	72.0	19.0	100.0	100.0	1210.0	147.0	634.0	186.9
Mean	39.9	3.9	58.8	72.1	134.7	17.0	83.8	13.3
Std Dev.	9.9	3.9	16.2	19.6	176.2	22.0	113.2	22.5
Skewness	0.3	2.3	-0.2	-0.6	2.4	2.6	1.9	2.7
Kurtosis	-0.1	5.0	-0.1	-0.4	7.0	8.4	4.2	10.9

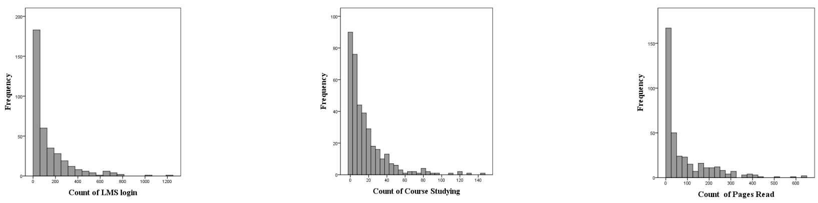

Figure 1: Frequency histogram of engagement data (LMS login, course study, pages read)

Negative Binomial Model: An Over-Dispersion Poisson Regression

Negative binomial regression is known as a log-linear model, with the dependent variable, a count variable, has a Poisson distribution. The logarithm of its expected value can be modeled by a linear combination of the predictors. Over-dispersion is occurred when the problems of excess zeros in which a subgroup of respondents who would never display the behavior are included in the sample. In this study, the regression analyses were performed for the three models, in which each included a different dependent variable as follows: Y1 is the counts of LMS login, Y2 is the counts of course studying, and Y3 is the counts of pages read, respectively. Age, gender, level of college years, time spent in e-learning, the midterm test scores, and the final exam scores served as the predictors.

Results and Discussions

In Table 2, the parameter estimates shows the negative binomial regression coefficients for each of the predictor variables along with their standard errors. Across the three models, the predictor variables, the midterm scores, the final scores, and time spent in e-learning, are statistically significant. It explained that, for instance, each one-unit increase on midterm scores, the expected log count of the Y1, Y2, Y3 increase by 0.01 times, 0.014 times, and 0.011 times respectively. The midterm scores have a stronger affect (B =0.014) on the counts of course studying (i.e., Y2) than Y1 and Y3. Each one-unit increase on time spent in e-learning, the expected logarithm of the pages read (i.e., Y3) increase by 0.044 pages. The categorical variable Level=1st year has a coefficient of -1.195, which has statistically significant effect only on Y1. As compared to senior students, the count of LMS login for the freshman was less. The time spent in e-learning for the male students is decreased by 0.274 hours, less than the female. The LMS login can be well predicted by the factors of age, midterm, final exam, time spent, and the levels of the college years. The college years may affect only on the frequency of using the LMS. Test scores and time spent in e-learning can be well used for predicting the students' online engagement.

Table 2: Parameter estimates of negative binomial regression by three dependent variables

Parameters N=364	df	B of Y1	Std Error	Wald Chi-Square	B of Y2	Std Error	Wald Chi-Square	B of Y3	Std Error	Wald Chi-Square
(Intercept)	1	2.845	0.321	78.42***	0.296	0.352	0.71	1.330	0.334	15.81***
Gender=male	1	0.007	0.108	0.00	-0.001	0.112	0.00	-0.274	0.109	6.38**
Gender=female		0			0			0		
Level=1st year	1	-1.195	0.150	63.20***	-0.068	0.160	0.18	0.052	0.154	0.11
Level=2nd~3rd yr	1	-0.161	0.122	1.74	-0.089	0.131	0.46	-0.047	0.124	0.14
Level=4th year		0			0			0		
Age	1	0.018	0.006	9.56**	0.010	0.006	2.70	0.023	0.006	14.31***
Midterm scores	1	0.010	0.004	7.10**	0.014	0.004	13.16***	0.011	0.004	7.47**
Final scores	1	0.007	0.003	4.92*	0.009	0.004	6.80**	0.007	0.003	4.92*
Time spent	1	0.022	0.003	53.25***	0.027	0.003	75.52***	0.044	0.004	152.90***

Note: *** stands for ρ -value less than the significance level at 0.001; ** : 0.01; *: 0.05.

Conclusion

This study has analyzed the data about e-learners' engaged behaviors in use of a negative binomial regression. The data collected by LMS in the case study shows that the effects of the predictors, such as age, test scores, and time spent on the counts variables. The findings provide a good fit of the new method and the count data types.

References

[1] Conrad, D. (2002). Engagement, Excitement, Anxiety, and Fear: Learners' Experiences of Starting an Online Course. American Journal of Distance Education, 16(4), 205-226.

[2] Hiltz S. R. (1994). Learning Together Online. Lawrence Erlbaum Association. New Jersey.

[3] Kirkpatrick, D. (1994). Evaluating Training programs. Berrett Koehler, San Francisco, CA.

[4] Robinson, C. & Hullinger, H. (2010). New Benchmarks in Higher Education: Student Engagement in online learning. Journal of Education for Business, 12,3, 100-120.

[5] Thurmond, V., Wambach, K., & Connors H. (2002). Evaluation of student satisfaction: Determining the impact of a Web-based environment by controlling for student characteristics. The American Journal of Distance Education, 60, 169–189.