Study on the comprehension process of university students using time-series analysis

Ayako OHSHIRO†*
a.ohshiro@okiu.ac.jp
†Department of Business Administration Okinawa International University
Ginowan City, 0kinawa 901-2701, Japan
*Correspondence

Summary
With the recent advances in information and communication technology, online management of students' learning data has become the norm. Research on learning analysis that predicts the near future (in a few years) of students' careers using machine learning methods and state transition models has been widely conducted. It is important for educators to evaluate the comprehension stability of students to prevent a decrease in their comprehension rate and dropouts in the class. In this study, we measured the comprehension process of university students in different types of lectures. Herein, we report on the results of data analysis using time series and data statistics, and consider several educational approaches.
Key words:
Time series, Data analysis, learning process

1. Introduction

Ever since the paradigm shift in the education system in the late 1980s, the rate of knowledge transmission in students has been evaluated using test scores, and the rate of knowledge accumulation has been evaluated using learning reports. With the advancement in information and communication technology, online management of students' learning data has become the norm. Research on learning analytics (LA) and educational data mining (EDM), for improving education quality by analyzing the learning data, has been widely conducted in recent years. Furthermore, most college lectures have migrated to online platforms in response to the novel coronavirus pandemic, thus significantly increasing the opportunities to obtain students' learning data. The knowledge accumulation-type lecture helps students evaluate themselves in each lecture with a report; thus, comprehension can be visualized between educators and students. It can help students in real time overcome their insufficient ability for comprehension and educators overcome their insufficient ability to conduct lectures.

The analysis of time-series data, which changes over time, has increased in the field of stock quote prediction [1], restaurant sales [2], precipitation [3][4], brain waves, and heart rate [5]. In addition to typical linear prediction methods, such as vector autoregression (VAR) and auto regressive integrated moving average (ARIMA), and machine learning methods, such as long short-term memory, are applied. The same methods are used in the field of education. Studies that analyze students' learning data [6][7] as time-series data predict their career using a state transition model, detect the students who are likely to drop out of school, and predict their academic grade using machine learning methods [8]. Kondo et al. [9] proposed a learning process model based on a Bayesian network that uses student data comprising submission of reports, attendance rate, and credit acquisition. Such learning data can have an impact on work performance after graduation in the long term, as it is evaluated by the accumulation of learning. The same is true for the short term (i.e., in college) where students attend 15–16 lectures per subject. Students' comprehension of the first lecture can affect the next lecture. This model proves helpful to educators in encouraging students to improve by understanding the degree of the effect in advance after each lecture. For example, educators noticed that an increase in students' comprehension helps prevent an increase in the dropouts by encouraging them to improve.

The aim of learning should be recognized by both educators and students to achieve quality education in college. The first lecture should be an orientation that includes a brief description of the syllabus, but it is a good information for lectures that educators understand of students' comprehension.

In our previous study [10], we focused on standardizing educational models by using data such as changes in students' perception before and after attending a training program. Moreover, an investigation was conducted on college students' recognition ability of the process of an information literacy educational program; it was concluded that there was a significant difference among the upper and lower grades, learning categories, and educators and students [11]. In this study, we focused on the process of students' comprehension of each lecture and significant differences in their academic grades (year of study).

In this study, first, we note college students' comprehension of each lecture and the difference in their

Manuscript received August 5, 2021
Manuscript revised August 20, 2021
https://doi.org/10.22937/IJCSNS.2021.21.8.23

academic grades. Second, we conclude the impact of the students' comprehension of the previous lecture on the subsequent lectures using a time-series analysis method.

2. ARIMA process

One of the characteristics of time-series data is the relationship between the past and present values. The auto regression (AR) process (defined by the correlation between the past and present values), moving average (MA) process (defined by their average), and ARIMA process [12][13] (integrated the AR and MA processes) were the proposed traditional time-series analysis methods. In these methods, a prediction model is constructed at point *t+1* using *t* or *t-d*. For example, the prediction model is indicated by the AR process using the number of lagged p, and the MA process using the number of lagged q, by using parameters α and β, as follows:

$$x_t = c + \varepsilon_t + \sum_{i=1}^{p} \alpha_i x_{t-1} + \sum_{j=1}^{q} \beta_j e_{t-j},$$

where x_t is a stationary variable, c is a constant, α_i is the autocorrelation coefficients at lags, and β_j is the weights applied to the current and prior values of the stochastic terms in the time series. The term e_{t-j} is assumed as Gaussian white noise series with mean zero and variance σ_ϵ^2.

3. Analysis of students' comprehension process

In this section, the process of surveying students' comprehension is explained. The subject types surveyed in this analysis can be classified into: classroom lecture type and hands-on practice type as a configuration and common type and specialized type as a subject. The common subjects are available to students of all grades and faculties, and the specialized subjects are limited to certain grades and faculties. Herein, we observed the process of students' comprehension by configuration type and subject type. The students were requested to give a score (using a 5-step scoring system of 20% or less, 40%, 60%, 80%, and 100%) to each lecture, based on their comprehension of the lecture.

First, the survey results were analyzed by observing basic statistical values of the comprehension scores from the collected data. In addition, based on a previous study [11], awareness change in students occurs from the first-year (early post-high-school period) to the second-year (students who have had time to adjust to college life). Thus, we observed the difference in their comprehension level based on their grades by classifying the students into "lower" and "upper" academic grades.

Second, we treated the score data as time-series data, which comprised 15 lectures, and concluded the effect of students' comprehension of the previous lecture on the subsequent lecture using the ARIMA model.

The procedure of administering the survey and the analysis of the students' comprehension process are described as follows:

Step 1 Students evaluate each lecture based their comprehension level, using a 5-step scale scoring system.

Step 2 The scores are obtained as time-series data.

Step 3 The statistical values, such as *average, median, maximum*, *minimum, standard deviation*, and *cumulative* comprehension level, are calculated for each lecture.

Step 4 The differences in their comprehension level are compared based on their grades by using analysis of variance (ANOVA)

Step 5 The effect of comprehension level of the previous lecture on the subsequent lectures is observed using the ARIMA model applied to time-series data.

Here, Steps 3 and 4 correspond to the procedure of "analyzing differences by years of study," and Step 5 corresponds to the procedure of "analyzing previous lecture comprehension level's effect on subsequent lecture."

4. Collected Data

A survey for five classes was conducted in 2020. In this investigation, 368 time-series data were collected, wherein $Subject_1$ and $Subject_2$ are common subjects of lecture type, $Subject_3$ and $Subject_4$ are specialized subjects of lecture type, and $Subject_5$ is a specialized subject of practical type. The differences in the common subject classes, which are $Subject_1$ and $Subject_2$, were analyzed based on the year of study. Details of the collected data comprised lecture type and number of students in each class, as shown in Table 1.

Table 1: Details of each class

	Lecture or Practical	Common or Specialized	Number of students
Subject 1	Lecture	Common	81
Subject 2	Lecture	Common	39
Subject 3	Lecture	Specialized	137
Subject 4	Lecture	Specialized	101
Subject 5	Practical	Specialized	10

5. Results and discussion

In this section, we discuss the process of students' comprehension from multiple perspectives using the results of the analysis. The statistical values for each class are shown in Fig. 1. The horizontal axis indicates the number of lectures, and the vertical axis indicates the comprehension level.

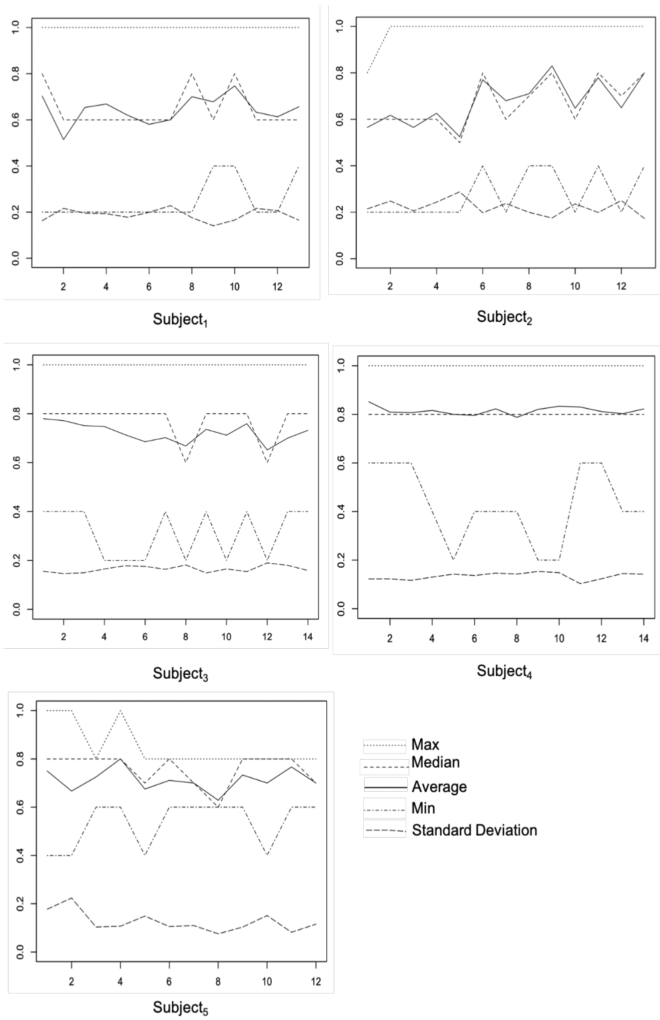

Fig. 1 Statistical values of the comprehension level for each class

As shown in Fig. 1, the average and median values of *subject$_1$* vary between 0.5 and 0.8, and the minimum value finally increases to 0.4. *Subject$_2$* has a larger standard deviation than Subject1, but the average, median, and minimum values increased while repeating the increase and decrease. The average and median values of *subject$_3$* and subject4 vary between 0.6 and 0.8 and has a considerable increment in the minimum value. As *subject$_3$*, *subject$_4$*, and *subject$_5$* focus on specialized content, we can conclude that the variance of students' previous knowledge may be reflected. The average, median, and maximum values of *subject$_5$*, which is a specialized subject of practical type, tend to decrease. This is probably due to the increasing complexity of the lecture content by the end of the class. The standard deviation values changed more in common subjects (*subject$_1$* and *subject$_2$*) than in specialized subjects (*subject$_3$* *subject$_4$* and *subject$_5$*). This may be the result of variation in students' specialized knowledge or years of study.

The possibility of differences in the comprehension level depending on the years of study concerning common subjects (*subject$_1$* and *subject$_2$*) was confirmed by ANOVA, as shown in Table 2. The processes of cumulative comprehension are shown in Figs. 2 and 3. The horizontal axis indicates the number of lectures, and the vertical axis indicates the cumulative comprehension level. The red and black lines indicate the years of study classified as "upper" and "lower," respectively.

Table 2: *p*-value in ANOVA for each common class by academic grade

Subject	Subject$_1$	Subject$_2$
p-value	3.323e-13	1.102e-13

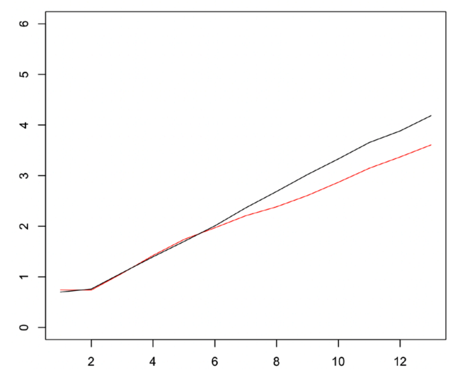

Fig. 2 Difference in the cumulative comprehension level depending on the years of study in *subject$_1$*

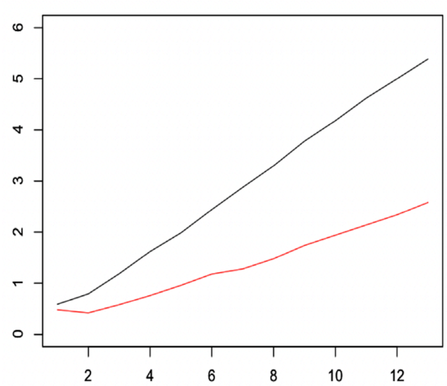

Fig. 3 Difference in cumulative comprehension level depending on the years of study in *subject$_2$*

As the *p*-values of both common subjects are significantly small, the comprehension level differs significantly by academic year. Moreover, as the lectures progressed, differences in cumulative comprehension level between academic grades classified as “upper” and “lower” increased and eventually, the “lower” exceeded the “upper.” We can conclude that it is desirable to take a common subject in the lower grades.

Finally, the effects of the comprehension level of the previous lecture on the subsequent lectures were confirmed using the ARIMA model, as shown in Table 3. The BIC was applied as an accuracy evaluation index.

Table 3: Comprehension affect between lectures with the ARIMA model

	Model	BIC
$Subject_1$	$x_t = 0.6436$	-32
$Subject_2$	$x_t = -0.7672x_{t-1}$	-20.85
$Subject_3$	$x_t = -0.6059x_{t-1}$	-40.51
$Subject_4$	$x_t = 0.8150$	-70.46
$Subject_5$	$x_t = 0.7130$	-35.75

We can find that the comprehension level of the previous lecture affected the comprehension level of the next subsequent lecture in $subject_1$, $subject_4$, and $subject_5$. In other words, almost all previous lecture’s comprehension levels had less effect on the next lecture’s comprehension. Similarly, the comprehension level of the previous lectures affects the comprehension level of the next lecture in $Subject_2$ and $Subject_3$ with negative coefficient values. This implies that previous lectures’ comprehension levels tend to easily decrease, and the result applies to the learning pyramid model, as shown in Fig. 4. The learning pyramid model [14] shows a 5% comprehension stability rate over time in the class of lecture type.

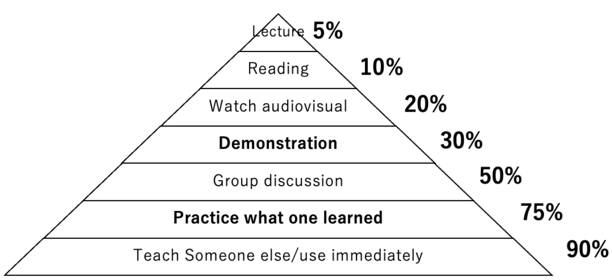

Fig. 4 Each comprehension stability with the learning pyramid

Therefore, the addition of “demonstration” or “practice” that one learned in the learning process can lead to an improvement in the comprehension stability such that it could be maintained at 30%–75%, as shown in Fig. 4, thereby preventing a decrease in comprehension, and dropouts in college.

6. Conclusion

To help improve comprehension stability and prevent dropouts from classes in college students, a survey was conducted to determine their comprehension process. In particular, the questionnaire item was students’ comprehension level at each lecture for several types of classes. The data analysis results showed that the comprehension levels differ depending on the type of class, type of subject, and academic grade. In addition, it is possible to improve students’ comprehension levels by combining various lecture styles. Based on the insights gained from this study, we hope to provide good education with appropriate awareness of classes.

References

[1] K. Nakagawa, M. Imamura, K. Yoshida, “Stock Price Prediction Using Similarity of Stock Price Fluctuation Patterns”, The 31st Annual Conference of the Japanese Society for Artificial Intelligence, 2D1-1, p.1-p.4 2017

[2] T. Iha, I. Kawamitsu, A. Ohshiro, and M. Nakamura, “An LSTM- based multivariate time series predicting method for number of restaurant customers in tourism resorts”, Proceedings of International Technical Conference on Circuits/Systems, Computer and Communications, p.1-p.4, 2021

[3] A. Luis, A. Diaz-Robles, C. Juan, A. Ortega , S. Joshua, B. Fu, D. Gregory, B. Reed , C. Judith , C. Chow ,G. John, C. Watson, A. Juan, A. Moncada-Herrera, “A hybrid ARIMA and artificial neural networks model to forecast particulate matter in urban areas: The case of Temuco, Chile”, Atmospheric Environment, 42 (35), p.8331-p.8340, 2008

[4] M. Małgorzata, I. Malinowska, M. Gos, and J. Krzyszczak, “Forecasting daily meteorological time series using ARIMA and regression models”, International Agrophysics, 32 (2), p.253-p.264, 2018

[5] J.L. Bernal, S. Cummins, A. Gasparrini, “Interrupted time series regression for the evaluation of public health interventions: a tutorial”, International Journal of Epidemiology, 46(1), p.348-p.355, 2017

[6] Y. Okada, T. Torii, "Process model of educational information management in institutional research for teaching and learning", 42 (4), p.313-p.322, 2019.

[7] G. Siemens, and R.S.D. Baker, “Learning analytics and educational data mining: Towards communication and collaboration”, Proceedings of the 2nd International Conference on Learning Analytics and Knowledge, p.252-p.254, 2012.

[8] N. Kondo, “Utilization of predictive models in education and learning”, Japanese Society for Information and Systems in Education, 37 (2), p.93-p.105, 2020

[9] N. Kondo, T. Hatanaka, “Modeling of learning process based on Bayesian network”, Japan Society for Educational Technology, 41 (3), p.271-p.281, 2018

[10] A. Ohshiro, S. Ueda, “Factors extraction for clinical research educational model and their relation”, IJCSNS International Journal of Computer Science and Network Security, 19 (3), p.87-p.91, 2019

[11] A. Ohshiro, "A survey of student's perception for information literacy", IJCSNS International Journal of Computer Science and Network Security, 20 (9) p.50-p.54, 2020
[12] G.E. Box, G.M. Jenkins, "Time series analysis: Forecasting and control", San Francisco: Holden-Day, 1970.
[13] D.F. Findley, B.C. Monsell, W.R. Bell, M.C Otto, and B.C. Chen, "New capabilities and methods of the X-12-ARIMA Seasonal Adjustment Program Findley", Journal of Business and Economic Statistics, 16, p.127-p.177, 1998
[14] K. Letrud, "A rebuttal of NTL Institute's learning pyramid", Education, 133 (1), p.117–p.124, 2012

Ayako Ohshiro received her B.S. and M.S. degrees from the University of the Ryukyus in 2009 and 2011, respectively. She completed her Ph.D. from the University of Ryukyus in 2017 and worked as an assistant professor at the University of the Ryukyus from 2015 to 2020. She has been a lecturer at Okinawa International University since 2020. Her research interests are data analysis in the field of intelligent information engineering.