

# Log Data Analysis of Learning Histories in an e-Learning Course

Koichiro Ishikawa, Masako Furukawa, Yutaka Saito, Takahiro Naito, and Mamiko Shichida

**Abstract**—Unfortunately, it has not yet been established fully the method how to analyze a log data set of learning histories obtained in an e-learning course. On the other hand, the researches in complex systems field report that patterns or regularities, such as power-law distribution, in higher or global level can be observed in many areas e.g., earthquakes, links between Web pages, sales result of books, etc. In this paper, we show that such an undisclosed macroscopic pattern or regularity in global level can emerge from the collective behaviors by learners in an e-learning course. We expect that such the patterns or regularities might be able to be employed as the useful information to understand learners' behaviors well and to make efficient learning courses.

**Index Terms**—E-learning, power-law distribution, complex system, learning management system (LMS).

## I. INTRODUCTION

In today's information-oriented society, e-learning becomes to play the indispensable role for education, learning, and training. One of the advantages of e-learning, especially for instructors and trainers, is that detailed data of learning history can be obtained easily under the e-learning environments.

In many e-learning courses, computer software called Learning Management Systems (hereafter referred as LMS) is used when learning materials are broadcasted (e.g., [1]). LMS can collect and record learning history data, which is executed automatically with no effort of instructors [2].

Though a large numbers of learning history data obtained from a large number of learners potentially useful for e.g., making the e-learning course effective, *et al.* The analysis of data does not proceed enough in many e-learning courses. One reason may be the difficulty of analyzing the data. The learning behaviors of only one learner can be complex to be understood. It is because his/her actions are decided with particular personality and are easy to be changed according to contexts.

It becomes more problematic when learning history data of many learners are obtained. Wide variety of behaviors by many learners may increase the difficulty of analysis. Moreover, huge amount of data, even if obtained with almost no effort by using LMS, make themselves difficult to be handled. It is because the amounts of data tend to hide which

data or feature is important. This difficulty is a common problem as data mining for a "big data" set.

On the other hand, an e-learning course can be understood as a complex system because, as explained fully in Section II-A, it has almost all of the eight key components of Complexity pointed out by Johnson [3]. From the viewpoint of the researches in complex systems field, a simple mechanism could emerge complex appearance. For example, [4], [5] reported that, through the simple mechanism, small differences among initial value sets can emerge wide variety of results.

As concerned with learning history data, there are some researches try to investigate a regularity or pattern in the higher or global level. For example, [2] is focuses on whether or not a power-law distribution is limiting distribution when a set of data on learning behaviors in an e-learning course is given.

It is hopeful because, if the disclosed regularity can be explained the underlying simple mechanism, it could lead to an understanding how the variety of learners' behaviors emerged. Thus, the knowledge might be acquired how the learning history data should be analyzed.

In this paper, we report that the same regularity as in [2] is observed again in the log data of learning history obtained in another e-learning course. We insist here that it can be an evidence that imply the macroscopic regularity, or pattern, in the higher or global level emerge from the corrective behaviors of a learner group in total.

## II. PREMISE KNOWLEDGE

### A. Complex Systems

Emergence is one of the research topics that attract attention of researchers in the field of complex systems (e.g., [6]). For example, under a self-organization process, complex structures at the higher level emerge solely from interactions among lower-level components without top-down controls or commands [7], [8].

Johnson pointed out that eight key components of a Complexity [3] as follows;

- 1) The system contains a collection of many interacting objects or "agents".
- 2) These objects' behavior is affected by memory or "feedback".
- 3) The objects can adapt their strategies according to their history, thus agents has autonomy.
- 4) The system is typically "open".

Then, he explains that the resulting system, i.e., complex system, will show the following behaviors as characteristics of Complexity;

- 1) The system appears to be "alive", i.e., the adaptation of

Manuscript received April 15, 2013; revised July 16, 2013.

K. Ishikawa is with the Chiba Institute of Technology, Chiba, 275-0023, Japan (e-mail: koichiro.ishikawa@it-chiba.ac.jp).

M. Furukawa, Y. Saito, and T. Naito are with the University of Tsukuba, Ibaraki, 305-8500 Japan (e-mail: furukawa@interc.tsukuba.ac.jp, saito.yutaka.gf@interc.tsukuba.ac.jp, tnaito@interc.tsukuba.ac.jp).

M. Shichida is with the Shimane University, Shimane, 690-8504 Japan (e-mail: mshichida@hirc.aoyama.ac.jp).

- agents shows as if the system is living.
- 2) The system exhibits emergent phenomena which are generally surprising, and may be extreme or unexpected results.
  - 3) The emergent phenomena typically arise in the absence of any sort of “invisible hand” or central controller, i.e. the system is self-organized without a line of command.
  - 4) The system shows a complicated mix of ordered and disordered behavior, i.e., dynamism or emergence of “pocket of orders” is observed.

Here, almost all the key components mentioned above are cases of an e-learning course. Learners can be regarded as agents with memories. They may interact and feedback each other. And if the evidence of the order or regularity is observed in learning behaviors, the learning course may be an example of complex system.

It is promising that the group of people may show some specific macroscopic patterns in global level, even though each individual acts in unexpected way (e.g., [9], [10]). The reason is that we may be able to regard a learner as a simple element or agent, as pointed out in [3]. Thus, it is no need to take account of his/her peculiarity such as personality.

If the varieties of leaning history data result difficulties to analyze them as mentioned in the previous section, the simple description may be helpful and learning history data collected in an e-learning course become really useful. Also, the simple description may lead to deep understanding for learning behaviors itself. Additionally, a simple description of group overall can lead to disclosure of processing rules that emerge global patterns.

### B. Power-Law Distributions

Power-law distribution is a probability distribution. It is reported that the power-law distributions are observed frequently under the self-organizing process in various areas [11]. Thus, a power-law distribution may be understood as one of such global structures. Many researches of complex systems are interested in both observing and validating that the limiting distribution of given a certain data is the power-law (e.g., [12], [13]). Power-law distributions are reported to be observed in e.g., the field of Internet, the numbers of Web pages that have certain number links to other Web pages, that of physics, the times of earthquakes with a certain magnitude occurred, and that of book sales, the number of book that are sold a certain number of copies, etc. [9]-[14].

One of the characteristics of power-law distribution is that the distributions have fat tail compared to e.g., a normal distribution [10]. Under the power-law distributions, the results can vary widely as compared to the results under a normal distribution. Thus, rather rare cases can be observed more frequently. For example, about book sales of Amazon.com [15], there are large numbers of book titles that are sold only one copy in one year, as well as there is a few book titles that are million sellers, both of which can be understood as rare cases. This fact becomes very popular as long-tail phenomena [16].

Granovetter shows that wide variety of results, i.e., how many people join a riot, comes out a rather simple experimental setting of simulation [3]. In the process of self-organization, local interactions between elements with less or no intelligence can results complex structures with

patterns or regularities. Such processes are researched on both real world phenomena and multi-agent simulations using computers (e.g., [5], [17]).

## III. RESULT

The Graduate School of Comprehensive Human Sciences, University Tsukuba (hereafter referred as UT), one of the national universities in Japan, has been opening e-learning courses for these four years. The aim of courses is to develop staffs on special medical areas. The courses are prepared with partly support by MEXT, Japanese government.

We analyzed the log data of learning histories that are obtained in the e-learning courses by UT that is opened in the academic year of 2012. From April 12012 through March 30 2013, 235,769 learning histories were recorded in the LMS server. We aggregated them into 57,132 contents viewing actions performed by 258 learners. The reason our aggregation process was needed is that one viewing action was divided and recorded as several partial histories, because e.g., the HTTP session was disconnected and reconnected instantaneously, etc.

Unfortunately, the LMS system operated by UT does not record the results whether or not a learner watched a viewing content throughout, i.e., from the start of the content to the end of it. Therefore, we could not help dealing such rather inappropriate data as time-diff values.

Without the exception mentioned here, we plotted basically after the technique proposed in [2] (See Appendix A. for detailed explanation of the technique in Section IV). The result is shown in Fig. 1.

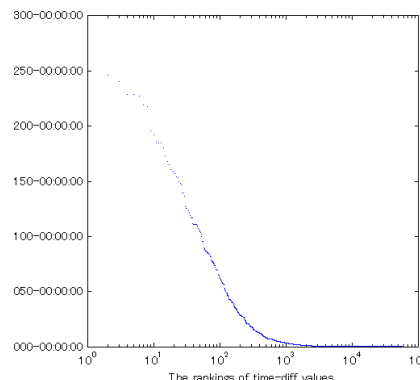


Fig. 1. Time-diff values and its rankings of the UT 2012 course.

Fig. 1 shows the result of plotting each time-diff value and its ranking. The time-diff values are contained the viewing actions that a learner did not watch a viewing content throughout, as mentioned above.

Y-axis of Fig. 1 shows time-diff values in normal scale. The X-axis shows the ranking of each time-diff values in all observed values plotted in log scale. These settings are after the previous research (in [2]), although, in almost all of such cases, it is popular to plot the data points in log-log scale (e.g., [16]).

Fig. 1 shows that the plots are almost in linear, especially from about 10 to 500 values of X-axis. The result is just like the one reported in [2]. Therefore, exponential values of Y-axis, not values of Y-axis themselves, and X-axis values observed in the UT 2012 course may follow power-law distribution as well as the result in [2].

#### IV. DISCUSSION

The result shown in Section III is meaningful because two e-learning courses, operated by UT and Aoyama Gakuin University (hereafter referred as AGU) which was reported in [2], are different from each other, as shown in Table I (See [18] for detailed information about AGU course).

TABLE I: THE DIFFERENCE BETWEEN UT COURSE AND THAT OF AGU

	UT Course	AGU Course
Property of learner	Graduate school students and paramedical staffs	Retired people and house Wives
Numbers of registered learner	Over 500 people	About 150 people
Course	Opened from 2011-	Opened 2008-2011
Term	Opened a year around	1.5 months course
Aim of course	Supplemental material for regular curriculum of Universities	Certification from authorized organization
Fields of contents	Medical and nursing knowledge	Psychology Coaching Laws (on Network), etc.
Number of viewing contents	Over 4000 materials	About 40 materials
Average length of viewing contents	About 12 min	About 10 min
Maximum length of a viewing content	About 100 min	About 17 min
Lerner support	No online support	Online support (by e-mentors)

The similarity shown in the result of two different courses can imply the macroscopic regularity or pattern in higher and global level. This regularity or pattern also can be emerged through some undisclosed mechanism that a learner group follows in total. The regularity of pattern means that time-diff values are scale-free and there is no typical value of time-diff (See [9] for more explanation). Thus, it is natural to observe an extreme large, or small, time-diff-value. Therefore, it may worthwhile to examine the existence of such a pattern of regularity in more e-learning courses other than above two.

On the other hand, Fig. 1 shows that the plots of the UT 2012 course are quickly close to 0 of Y-axis, especially data that is larger than about 500-th in rankings, i.e., X-axis. The reason may be that the UT 2012 course data contains viewing actions which learners did not watch the whole part of contents, as mentioned in the previous section. It can lead to smaller values of time-diff.

Unfortunately, the smaller values of time-diff can lead to the fact that data points plotted in linear is not many, i.e., about hundreds points, in Fig. 1. This problem is expected to be solved soon in the future because of two reasons as follows. First, log data of learning histories have been recorded continuously after 2012 course. The data set with that of newly recorded is expected to be plotted in linear in wider range of X-axis than the result of Fig. 1. Second, as explained in the next section, the new LMS system has already started working, which is expected to provide higher quality data.

Moreover, the LMS operated by UT allowed that a learner watch several different viewing contents in parallel at a time. Such cases was occurred when e.g., a learner access from several different browser software on one computer, etc. Thus, we ignored such data as far as we noticed. This may

cause the possibility that the final result is different from Fig. 1. However, we believe that difference is small and insignificant, because we focus on the tendency of global level that is less affected by the small changes of each value.

#### V. FUTURE WORKS

There remains several future works. As we insist in the previous section, researches of other e-learning courses than UT and AGU are indispensable. The reason and mechanism that emerge the result should be researched too. We also have to validate the adequateness of a power-law distribution as a limiting distribution fully, as well as we have to compare the results of two courses according to especially the gradients of two regression lines, etc.

Additionally, we have to consider how to use the result in order to make the learning course more effective, etc. The result may be useful information for effective learners support, i.e., mentoring under e-learning environments [19], [20].

About UT course, new LMS systems have already started working. The problem a learning action is divided and recorded as several histories is solved. Thus, the quality of data is expected to be higher, as well as the data handling process itself become simpler than ever. Moreover, it may be efficient to disconnect and reconnect HTTP sessions compulsory and automatically when some irregular access is detected.

In the previous section, we pointed out two problems on the LMS of UT. One of the problems is that LMS do not manage whether a learner watched a viewing content completely. Another problem is that a learner can watch viewing contents in parallel. These problems should also be solved in the future to obtain high quality data.

#### VI. CONCLUSION

In this paper, we report that the same macroscopic pattern or regularity in global level as reported in [2] is observed among the data obtained in another e-learning course. We also insist that the pattern or regularity emerges through collective behaviors of a learners group in total.

In the research field of education and educational technology, the characteristic of learners or learners groups as complex systems is not fully researched yet. The approach of this research shows possibly the underlying relationship between education, educational technology, and complex systems. Moreover the result may lead to understanding of learners' behaviors and might be able to be employed as the effective information to support learners efficiently, etc.

#### APPENDIX

##### A. Time-Diff Values

Fig. 2, originally published in [2], shows how to calculate time-diff values.

Reference [2] reported that, in seven courses they opened, participants pushed the play (start viewing) button 5,077 times in total and they finished watching each material throughout for the first time. There were 3,187 times out of 5,077 that participants pushed the button after more than the

average viewing time (about 15 minutes) when the participant had pushed the button last time. They referred to this time lag between sequential two button pushes as the time-diff.

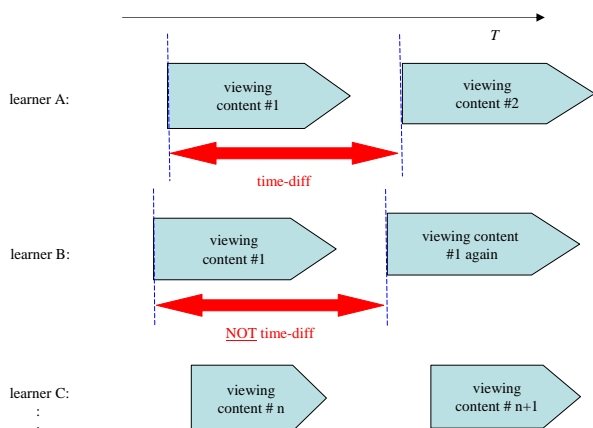


Fig. 2. The calculation of time-diff values. Originally published in [2].

As we discussed in Section IV, in [2], they ignored the viewing histories that do not continue from the start of a content through the end of it. This is a difference between the time-diff values in this paper and that in [2].

#### ACKNOWLEDGMENT

This research was partly supported by Grant-in-Aid for Scientific Research (C) 23501170, MEXT, Japanese government.

The learning history data was obtained in the e-learning courses opened by University of Tsukuba (UT). The first author thanks the people, especially the members of UT, who has been prepared, constructed, operated and concerned with the course.

#### REFERENCES

[1] Moodle. [Online]. Available: <http://moodle.org/>

[2] K. Ishikawa, M. Furukawa, M. Shichida, and Y. Saito, "A macroscopic pattern at global level shown in learners' behaviors," in *Proc. the 2nd Int'l Conf. on Behavioral, Cognitive and Psychological Sciences (BCPS 2011)*, ISBN 978-981-07-0624-1, 2011, pp. 62-66.

[3] N. Johnson, *Simply Complexity: A Clear Guide to Complexity Theory*, Oneworld, 2007.

[4] M. Granovetter, "Threshold models of collective behavior," *American Journal of Sociology*, vol. 83, no. 6, pp. 1420-1443, 1978.

[5] K. Ishikawa, B. Chu, A. Sakurai, and H. Matsukawa (D.-b. Cao), "An analysis of MML agent simulation under uniform distribution of gain parameter," *Information: An International Interdisciplinary Journal*, vol. 16, no. 7(A), pp. 4693-4704, ISSN: 1343-4500, 2013.

[6] S. Johnson, *Emergence: The Connected Lives of Ants, Brains, Cities, and Software*, Scribner, 2002.

[7] S. Camazine, J. L. Deneubourg, N. R. Franks, J. Sneyd, G. Theraulaz, and E. Bonabeau, *Self-Organization in biological systems*, Princeton University Press, 2003.

[8] K. Ishikawa, Y. Shinozawa, and A. Sakurai, "Self-Organization and Aggregation of Knowledge," in *Self Organizing Maps - Applications and Novel Algorithm Design*, J. I. Mwasigi, Ed., InTech, ISBN 978-953-307-546-4, 2011, chapter 9, pp. 143-172.

[9] M. Buchanan, *Ubiquity: The Science of History or Why the World is Simpler than We Think*, Crown Publishers, 2000.

[10] M. Buchanan, *The Social Atom: Why the Rich Get Richer, Cheaters Get Caught, and Your Neighbor Usually Looks Like You*, Bloomsbury, 2007.

[11] A. L. Barabási, *Linked: The New Science of Networks*, Perseus Books Group, 2002.

[12] A. Clauset, C. R. Shalizi, and M. E. J. Newman, "Power-Law distributions in empirical data," *SIAM Review*, vol. 51, no. 4, pp. 661-703, 2009.

[13] A. Clauset. Power-Law distributions in empirical data. [Online]. Available: <http://tuvalu.santafe.edu/~aaronc/powerlaws/>.

[14] M. Mitzenmacher, "A brief history of generative models for power law and lognormal distributions," *Internet Mathematics*, vol. 1, no. 2, pp. 226-251, 2004.

[15] Amazon. [Online]. Available: <http://www.amazon.com>.

[16] C. Anderson, *The Long Tail: Why the Future of Business is Selling Less of More*, Hyperion, 2006.

[17] E. Bonabeau, G. Theraulaz, and M. Dorigo, *Swarm Intelligence: From Natural to Artificial Systems*, Oxford University Press, 1991.

[18] K. Ishikawa, M. Shichida, M. Yuyama, Y. Saitou, and K. Tamaki, "E-mentor development course," in *Proc. Int'l Sympo. And Conf. for Educational Media in School (ICoME 2009)*, 2009, pp. 87-92.

[19] K. Ishikawa, M. Shichida, and Y. Saito, "Toward efficient learner supports under the mobile leaning environments," in *Proc. Int'l Conf. on Mobile and Context Learning (mLearn 2010)*, 2010, pp. 384-385.

[20] K. Ishikawa, M. Furukawa, M. Shichida, and Y. Saito, "E-Mentor autonomy: Learners' support in the next generation," in *Proc. the 2nd Annual Int'l Conf. on Education & e-Learning (EeL 2012)*, pp. 127-131, DOI: 10.5176/2251-1814\_EeL12.65, 2012.



**Koichiro Ishikawa** graduated from the Faculty of Economics at Yamaguchi University in 1989 and joined NEC Corporation. He received a master's and Ph. D. degree from the School of Knowledge Science at the Japan Advanced Institute of Science and Technology in 2000 and 2005 respectively. In 2005 he became a researcher at Tokyo University of Agriculture and Technology. In 2011 he became a researcher at University of Tsukuba. He has been an associate professor at the Chiba Institute of Technology since 2012. He also is a visiting associate professor at Keio University, and a visiting researcher of the Human Innovation Research Center, Aoyama Gakuin University. Dr. Ishikawa specializes in artificial neural networks, machine learning and educational technology. He is a member of JSISE and JSPM.



**Masako Furukawa** received her B.A. in literature 2000 and M.A. in literature 2002 at Showa Women's University Japan. Education at the Graduate University for Advanced Studies (SOKENDAI) Japan, 2002-2008. She works at International Student Center, University of Tsukuba Japan as a researcher. She is interested in researching and developing e-learning for Japanese Language Learning.



**Yutaka Saito** was born in Yokohama Japan in September 1973. After graduation from Bunkyo University in 1996, He received a master degree by graduate of Graduate School of International Politics, Economics and Communication at Aoyama Gakuin University Japan in 2002. He became an assistant Professor at Aoyama Gakuin University Institute of Human Innovation Research Center in 2008 and Aoyama Gakuin University Institute of Information Science in 2011. He became a visiting researcher at Tsukuba University faculty of medicine in 2013. He studies educational technology now.



**Takahiro Naito**, 2001 Graduated from Japan Advanced Institute of Science and Technology, Graduate School of Knowledge Science. 2007 Graduated from Tokyo Medical and Dental University, Graduate School of Biomedical Science, Biomedical Science PhD Program. 2008-2009 Researcher, Graduate School of Comprehensive Human Sciences, University of Tsukuba. 2009 Assistant Professor, Faculty of Medicine, University of Tsukuba.



**Mamiko Shichida** was born in Tokyo Japan in July 1970. After graduation from the Toyo University in 1991, she joined Nihon Univ. Sakuragaoka High School as a teacher on Japanese Literature. She received a master's degree by research in the ancient literature of Japan at Saitama University Japan in 2004. She received a degree of Doctor of Literature at Graduate University for Advanced Studies (SOKENDAI) Japan in 2009. She became a visiting researcher at Aoyama Gakuin University Institute of Human Innovation Research Center in 2009, and associate Professor at Shimane University in 2013. She studies education now.