Learning Feature Analysis for Quality Improvement of Web-Based Teaching Materials Using Mouse Cursor Tracking

Mizuho Ikeda
School of Science and Technology, Kwansei Gakuin University
2-1, Gakuen, Sanda, Hyogo, 669-1337 Japan
gpikeda@kwansei.ac.jp

ABSTRACT

In blended learning, it is meaningful to obtain data on and understand the nature of the usage of web-based teaching materials by individual students and by whole classes in progress. For this purpose, a system called MCTA1, which can detect the behavior of students' mouse cursors on web-based teaching materials in real time and visualize the analysis results, was developed. MCTA1 provides functions that collect specific granularity as subsections and visualize the result of analysis in real time. Through the application of this system in real classes, particularly in programming practice lessons, it became clear that the behavior of the mouse cursor was an effective source of information for determining how individual students and whole classes actually use the web-based teaching materials. This paper describes an outline of how the MCTA1 system works. The analysis of students' usage under the actual conditions of each subsection is shown using a Lorenz curve and Gini coefficient. Additionally, the result of the analysis of the web browsing order route of each subsection is shown using Ward's clustering method.

Keywords

Mouse Cursor Tracking, Learning Feature, Quality Improvement, Lorenz Curve

1. INTRODUCTION

With the spread of education via electronic mediums , under various specialized fields, studies on the learning style [1], the effects of the education method utilized [2], and the operability of these systems have been actively conducted.

In order to improve the quality of the content of the web-based teaching materials, it is important to analyze exactly how the students use these teaching materials in their classes. In blended-type education, it is necessary to grasp information on the usage of web-based teaching materials under real conditions and determine the students' usage of these materials during actual classes in progress. In this situation, collecting the mouse cursor tracks is considered one of the effective means for determining their actual usage of the web-based teaching materials. Therefore, a system called MCTA1, which can detect the behavior of students' mouse cursors on web-based teaching materials in real time and visualize the analysis results, was developed. Since the system does not hang from the load that normally results from using special instruments on students, it can collect more precise data on the students' natural state. The quality of web-based teaching materials can be considered to consist of three components (curriculum composition, structure of web-based teaching materials, and lesson planning). The structured design of each quality component was drafted, and web-based teaching materials were constructed [3].

These materials have been used in blended learning sessions that use the e-learning system. The average number of members in a class is 25. Each session runs for 90 min, and each subject is performed over 14 sessions. The students who take the class belong to a department of liberal arts and are assumed not to have much knowledge of mathematical science and information science.

In this paper, a system that can detect the behavior of a student's mouse cursor is introduced. The action data from each subsection of the web-based teaching materials pages, which have been developed and implemented, were also discussed. Furthermore, the effect of analyzing the students' unique behavior via the segmentation granularity of the learning log is also described.

2. RELATED STUDIES

When LA research is roughly classified, it consists of the evaluation of the collection processes of (1) study data, (2) analysis, and (3) analysis result. From the point of view of detecting student behavior, learning the history data from the learning management system (LMS) as a collection of data in (1) above, the discovery of digital data, such as mouse cursor behavior [4]. facial actions [5], eye tracking [6] has been studied using surveys of self-reported data. Moreover, in order to estimate it as an analysis (2) of the above, or the above-mentioned analysis result (3), various visualization techniques have been proposed. In a research on the learning style and actions of a student, most of the experiments applied limited experimentation and evaluation.

Mizuho Ikeda "Learning Feature Analysis for Quality Improvement of Web-Based Teaching Materials Using Mouse Cursor Tracking" In: *The 12th International Conference on Educational Data Mining*, Michel Desmarais, Collin F. Lynch, Agathe Merceron, & Roger Nkambou (eds.) 2019, pp. 576 - 579

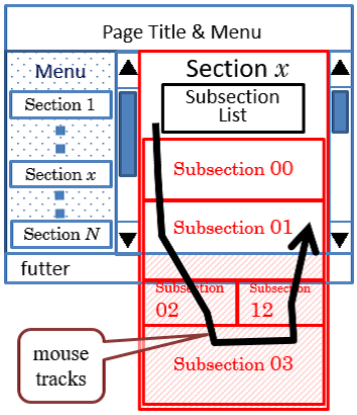


Figure 1: Web page composition and mouse cursor trackings.

These experiments were also performed in a restricted environment. There are few studies in real classes with conditions of continuous evaluation. Many analyses on the behavior of individual students have been conducted, but these studies often included only a few students and were limited by a lack of visualizations of group behavior.

When browsing a web page, the searcher tends to follow the mouse cursor with their eyes in order to easily click on the link of interest on the web page [7]. Based on this, the method of detecting the action of a mouse cursor is employed by this research. The system then detects whether the cursor would be in the position of a particular subsection based on the coordinate information built from the actions of many mouse cursors. The design and mounting were performed such that the automatic collection of continuous data could be realized, and a complete analysis of a student's features was conducted.

Section 3 describes the requirements for the functional enhancement of learning log collection, the system design of the action detection of a mouse cursor, and the mounting. Section 4 shows the system applied to a real classroom session and the effect of the feature analysis based on the learning log, subdivided according to the mouse cursor behavior data.

3. MECHANISM OF MCTA1

The web-based teaching materials used in real classes some pages consists of a menu, which contains the links to each section, and a section that is composed of several subsections.

At the top of each section, the subsection link is shown. Each indication range next to the subsection list is distinguished with a subsection number xy ($x = 0 \sim 9$, $y = 0 \sim 9$). Here, x shows the right and left domain, and y shows the top and bottom domain. Whenever the mouse cursor moves from one subsection to another, the numbers of the subsections that the cursor moves over in the process are sequentially collected automatically. The mechanism that collects both the subsection numbers and the time that a mouse cursor has been in existence has been developed.

Figure 1 is an example where the web-based teaching materials are comprised of five sections, and Subsection00 and a part of Subsection01 are displayed first. The shadowed area in the middle of Subsection01, Subsection02, Subsection12, and Subsection03 is shown. The arrow in Figure 1 is a trace of the mouse cursor tracks moved [Subsection00] → [Subsection01] → [Subsection02] → [Subsection03] → [Subsection12] → [Subsection01]. The y-axis direction is collected as the data "012321," and the x-axis direction is collected as "000010." At the same time, the time spent in each subsection is collected automatically.

The analysis and its visualization in real time are enabled by detecting the coordinate information, showing the action of a mouse cursor per subsection. The subsection of the middle position is only passed if the scroll bar is used and the trace data of the mouse can be collected sequentially. In addition, even if multiple windows such as when multiple tabs or separate windows are open are used alternately, the data can still be accurately collected.

4. VISUALIZING LEARNING FEATURE

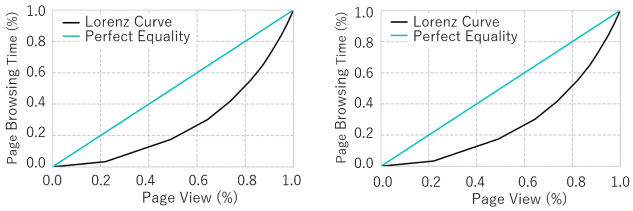
By using mouse cursor tracking data, the distribution difference of the browsing subsections is validated. The data also allows checking of whether the browsing distribution of the web-based teaching materials does not result in a difference when the unique features of each student are analyzed. Here, a Lorenz curve and the Gini coefficient are used to evaluate the inequality of page browsing as a statistical value. In fact, these are commonly used statistics to express inequality, such as with income in economic science. As the Gini coefficient approaches 0, the Lorenz curve approaches the complete equality line, and the equability becomes high. As the Gini coefficient increases, the Lorenz curve keeps away from the complete equality line, income disparity decreases, and the equability decreases. The Lorenz curve, the complete equality line, and the Gini coefficient were used to analyze the reference time of each subsection of the web-based teaching materials.

The possibility of referring to each subsection equally becomes high, so that the Gini coefficient calculated from the accumulation reference time of web-based teaching materials approaches 0. In a blended-type educational setup, the students are expected to follow the footsteps to the subsection which the teacher is explaining and likely to refer to the said subsection. Therefore, the closer the Gini coefficient is to 0, the higher the possibility that the students are wandering around to another subsection. The higher the Gini coefficient, the lower the equability of the subsection reference. One subsection is more likely to be intensively referenced.

To determine the students' usage situation, clustering was then performed on the patterns of the browsing orders of the subsections. Ward's hierarchical agglomerative clustering method was adopted because it tended to be easy to clearly classify into a cluster.

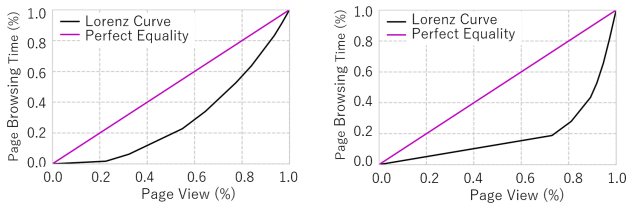
4.1 Visualization using Lorenz curve

A Lorenz curve, represented by the values along the y-axis, is the cumulative relative frequency of the browsing time



(a) Level 2. **(b)** Level 5.

Figure 2: Lorenz Curve (Level)



(a) Session 2. **(b)** Session 13.

Figure 3: Lorenz Curve (Session)

of subsections, with its x-axis being the cumulative relative frequency of the browsing number of subsections used to visualize the inequality of the browsing time distribution. First, the browsing distribution according to each student's level, i.e., degree of subject achievement, was compared.

Based on achievement in the applied subject, the five steps for gauging the student's levels were evaluated, and a group division was carried out. The level is a five-step evaluation based on the number of tasks achieved for 38 subjects.

The Lorenz curve for level 2 is shown in Figure 2(a), and for level 5 is shown in Figure 2(b). Here, level 5 implies the level with the highest degree of subject achievement. The Gini coefficient was 0.45 at level 2 and 0.44 at level 5; therefore, the difference was not so apparent. Next the browsing distribution according to each lesson was compared. The browsing time distributions of the contents of the web-based teaching materials for all students of the 2nd session and the 13th session are indicated in Figure 3(a) and Figure 3(b), respectively. The Gini coefficient for the 2nd session was 0.41 and for the 13th session was 0.60. Compared with the first session time, the Gini coefficient tends to increase. This demonstrates that the inequality of the browsing distribution becomes high. Therefore, it is shown that the possibility of perusing one's subsection is high toward the second half of a lesson.

4.2 Time series display of Gini coefficient

The time series transition of the Gini coefficient in Figure 5 is a graph in which the session extends in the x-axis direction and Gini coefficient is set to the y-axis direction. The lower the Gini coefficient, the higher the browsing homogeneity. This shows that various subsections are browsed uniformly. The higher the Gini coefficient, the higher the uniformity. This shows that a few subsections are browsed intensively.

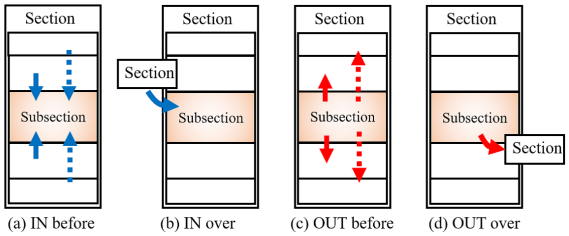


Figure 4: Patterns of browsing order.

The Gini coefficients of students from all levels except level 1 do not fluctuate very much. Particularly during the second half of the term (Sessions 8–14), the Gini coefficient of the levels 4 and 5 students maintain a relatively static state in a location for which it is high. In the first half of the term, the students browse not only the same subsections as the teacher, but other subsections as well. Therefore, the browsing distribution can encounter an inequality fluctuation for a short while. It later turns out that as the second half of the term progresses, the students become more concentrated in one subsection while browsing.

The level 2 students follow the teacher's browsing transition to the subsection, similar to what the level 5 students do in the first half of the term (Sessions 1–7). However, after Session 8, the homogeneity of the students' subsection browsing becomes much higher, and it seems to become clear that they browse various subsections, including past subsections, in addition to the subsections that the teacher browses.

In addition, the figure representing the time series of the Gini coefficient can be shown in real time. The teacher is able to confirm the non-homogeneity of the browsing distribution and adjust the speed of the teaching progress by returning to past subsections and lectures. Furthermore, it becomes useful in thinking about web-based material contents and the relocation of subsections.

4.3 Browsing order and route

Next, to clarify the students' features, representing their browsing behaviors, their browsing routes to each subsection were analyzed.

4.3.1 Patterns of browsing order

The students' mouse cursor tracks to each subsection are classified into the following four patterns. The pattern of where the mouse cursor came from and where it was moved to can be classified into the following four types (see Figure 4):

(a) **IN before :** Move from one subsection to the target subsection in the same section

(b) **IN over :** Move from one section to the target subsection

(c) **OUT before :** Move from the target subsection to another subsection in the same section

(d) **OUT over :** Move from the target subsection to another section

4.3.2 The graph of each cluster feature

According to these attributes, clustering is performed and a dendrogram is constructed by Ward's hierarchical agglom-

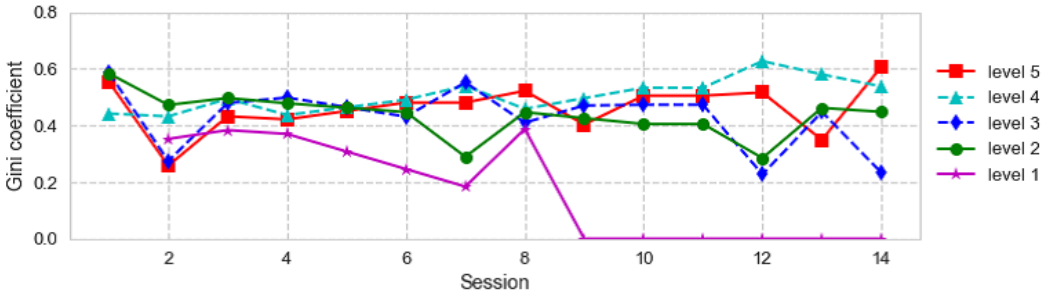


Figure 5: Transition of the Gini coefficient of each session (each student level).

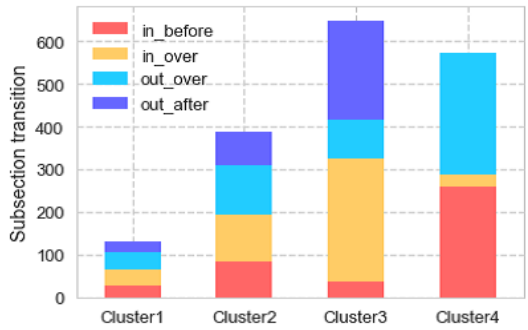


Figure 6: Characteristic of the clustering result.

erative clustering method. In Figure 6, when height is set to 750, number of clusters is set to four. To confirm the four clusters' features, stacked graphs were created. According to the list of subsections of each cluster, the following features were clarified:

Cluster1 : Students browse repeatedly through various subsections without following the composition order of the subsection. It occurs in the subsection concerned with the beginners' tasks in the first half of the term.

Cluster2 : Like in "Cluster1," students browse repeatedly through various subsections without following the composition order for the subsection. However, the subsection is browsed more times than in "Cluster1." It is also occurring in the subsection concerned with basic tasks in the first half of the term.

Cluster3 : Students transferred from the subsection without following the order of subsection composition and transferred to the next subsection according to a composition order. It is occurring in the subsection concerned with the high difficulty level tasks in the second half of the term.

Cluster4 : Students transferred from the subsection according to a composition order and transferred to the other subsection without following the order of subsection composition. Like in "Cluster3," it is occurring in the subsection concerned with the high difficulty level tasks in the second half of the term.

Classifying subsections seems to help with the reconstruction of the web-based teaching materials so that students' understanding can be increased in stages.

5. DISCUSSION

Visualization using Lorenz curve, the time series transition of the Gini coefficient and the graphs of the clustering methods can be leveraged for information to support the class as the lessons progress. In addition, it was found that the results of the clustering of students' transition route is useful information for reviewing reconstruction, rearrangement, and speed adjustment of class progress, as the quality model of web-based teaching material.

6. CONCLUSION AND FUTURE WORK

A system to collect mouse cursor behavior automatically has been designed and developed. In addition, it has been applied to an actual class. Learning logs have also been collected. Furthermore, a system that can visualize in real time the progress of a lesson using the learning log has been developed. It was found that it is possible to clearly capture a student's learning activities. A learning log of the subsection units has been accumulated every semester since the 2015. A quality evaluation model of web-based teaching materials needs to be established by analyzing additional learner features (e.g., the students' learning style) and applying them to a mathematical model.

7. REFERENCES

[1] T. A. Litzinger, S. Lee, J. C. Wise, and Richard M. Felder. A Psychometric Study of the Index of Learning Styles. In *Journal of Engineering Education.* 96(4), pages 309–319, 2007.

[2] N. Bos, B. Meijer and S. Brand-Gruwel. A network analysis of blended learning. In *EARLI SIG4*, 2016.

[3] M. Ikeda. A Development and Implementation Model for Evaluating Web-Based Teaching Materials. In *Research Report of JSET Conferences.* 2012(4), 49–54, 2012.

[4] J. Schneider, M. Weinmann, J. vom Brocke, and C. Schneider. Identifying Preferences Through Mouse Cursor Movements Preliminary Evidence. In *ECIS2017*, pages 2546–2556, 2017.

[5] A. K. Vail et al. The Affective Impact of Tutor Questions: Predicting Frustration and Engagement. In *EDM2016*, pages 247–254, 2016.

[6] S. Hutt, C. Mills, S. White, P. J. Donnelly, and S. K. D'Mello. The Eyes Have It. : Gaze-based Detection of Mind Wandering during Learning with an Intelligent Tutoring System. In *EDM2016*, pages 86–93, 2016.

[7] P. B.Kantor, E Boros, B. Melamed, V. Menkov, B. Shapira, and D. J. Neu. Capturing Human Intelligence in the Net. In *ACM 08/2000*, Vol.43, No.8, pages 112–115, 2000.