On-line Contents Analysis System for e-learning

Maomi Ueno
Nagaoka University of Technology
ueno@kjs.nagaokaut.ac.jp

Abstract

This paper proposes a new contents analysis method for e-learning by using response time data. The unique features of this paper are as follows:1)From Information Theoretic approach, the Gamma distribution is derived as a probability distribution of response time in the e-learning. 2) The two parameters in the Gamma distribution, α and β, are respectively interpreted as follows: The parameter α means "Complexity of the content (which means the numbers of simple understanding processes to understand or solve the content)" and the parameter β means "Easiness of the simple understanding process in the content". As a contents analysis method, this paper proposes "α - β chart ". 3) An online system by using the proposed method is introduced. Furthermore, this paper demonstrates some efficient points of the developed system.

1. Introduction

Recently, distance education by using e-Learning has become popular in actual educational situations. One of the advantages of the e-learning is that it is easy to get huge learning histories data which has been saved as log data in the e-Learning. In this case, it is important how we store this data or how we utilize this data[1][2]. One of ideas to utilize this huge data efficiently is application to evaluate the contents quality. However, almost utilizations in the standard LMSs(Learning Management Systems) are to show only averages and variances of test scores, learning time data, or the numbers of learning times. It is necessary to develop more efficient data analysis methods for the learning historical data and apply them to the design of LOM (Learning Object Metadata). The main ideas of this paper is to develop a new characteristics analysis method of e-learning contents by using a mathematical model[3] and the application to the learning or response time data in e-learning. This paper employs Gamma analysis method for response time data[3]. This was developed to analyze response time data from the computer testing data. The unique feature of this paper are as follows:

1. The interpretation of the parameter β, which was not discussed in [3], is derived from mathematical meaning.
2. Using the results of 1, this paper proposes a evaluation method of the content by using $\alpha - \beta$ chart.
3. An on-line contents analysis system by using this evaluation method is developed and it is demonstrated in this paper.

2. Mathematical Model

This paper provides a new evaluation method of the response time data obtained by the e-learning system. Here, let a mathematical probability distribution model for the data be derived from the maximizing Entropy model[3].

Now, let consider response time that learners response independently a simple task. When the expected time E and the minimum time t_0 is given, it is assumed that the entropy

$$H[f_S(t)] = \int_0^\infty f_S(t) \log f_S(t) dt \qquad (1)$$

is maximized.

Then, the following exponential distribution as a response time distribution is derived,

$$f_S(t) = \frac{1}{\tau} e^{-(1/\tau)t} \qquad (2)$$

where $\tau = E - t_0$.

The actual response time data (learning a content, or solving an item) can not be assumed as a simple task, then (2) is expanded to the following Gamma distribution given by integrating α exponential distributions,

$$f(t) = \frac{\beta(\beta t)^{\alpha-1} \exp(-\beta t)}{(\alpha-1)!} \qquad (3)$$

Proceedings of the IEEE International Conference on Advanced Learning Technologies (ICALT'04)
0-7695-2181-9/04 $20.00 © 2004 **IEEE**

As the results, the response time distribution $F(t)$ is given by

$$F(t) = \begin{cases} 0 & t < t_0 \\ \int_0 f(t)dt & t \geq t_0 \end{cases} \quad (4)$$

There are two parameters α and β in the model, they can be estimated respectively as follows:

$$\hat{\alpha} = \frac{\hat{\tau}^2}{\hat{\sigma}^2} \quad (5)$$

$$\hat{\beta} = \frac{\hat{\tau}}{\hat{\sigma}^2} \quad (6)$$

where the hat indicates the estimators of the parameters and $\hat{\sigma}^2$ is the estimated variance of the response time.

3. Parameters interpretations

3.1. Parameter α

The parameter α indicates the numbers of the simple tasks which is needed to understand the content or solve the item. When the value of α is larger, then the slope of the Gamma distribution becomes slower. Conversely, when the value of α is smaller, then the slope of the gamma distribution becomes sharper. As the results, the parameter α is interpreted as "Complexity of the content" to understand or solve[3].

3.2. Parameter β

The relationship between τ and α is given by

$$\frac{\tau}{\alpha} = \frac{\sigma^2}{\tau} = \frac{1}{\beta}$$

This is, the inverse of β indicates "the average time of the simple tasks in the content". From this, β can be interpreted as "Easiness of the simple understanding process in the content".

3.3. Contents characteristics analysis method by using $\alpha - \beta$ chart

This section introduces a contents characteristics method by using $\alpha - \beta$ chart. Now, let consider the standardized estimated parameters α^*, β^* .Then, the contents whose parameters in the $\alpha^* - \beta^*$ chart can be interpreted as shown in Figure 4.

4. On-line contents characteristics analysis system

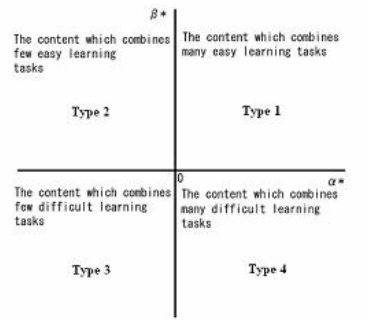

Figure 1. The contents characteristics analysis by using $\alpha - \beta$ chart

4.1. Outline

The author has developed an on-line characteristics analysis system which was included in our LMS mentioned before. The system is shown in Figure 2.The system renews the estimated parameters in the model as soon as a learner responses for a content or a test item. This frame shows subject name, the number of the students, lesson name, content name, the rate of learners' understanding and it's variance, Average response time and it's variance, the average number of the time which earners changed their answer and it's variance, and α and β .

The graph in the bottom of the figure shows the Gamma distribution. The broken line indicates the curve of actual data and the solid line indicates the theoretical curve.

4.2. Numerical Examples of response curves

The author has executed e-learning by using this system as an actual distance university class for one year. Now, the LMS has 78 courses and the total number of the participants are 461. Some examples of response curves are selected in the course "Computer" and demonstrated. The examples are divided into 4 Type categories.

(1) Type 1
The content corresponding to Figure 3 is a test item content located in Type 1. The figure means that this content is constructed by many simple tasks which do not need long time.

(2) Type 2
The content corresponding to Figure 4 is a animation content located in Type 2. The figure means that this content is constructed by few simple contents which do not need long time.

(3) Type 3
The content corresponding to Figure 5 is a image content located in Type 3. The figure means that this content is constructed by few simple contents which need long time.

(4) Type 4
The content corresponding to Figure 6 is a image content located in Type 4. The figure means that this content is constructed by many simple contents which need long time. This content needs much time to understand.

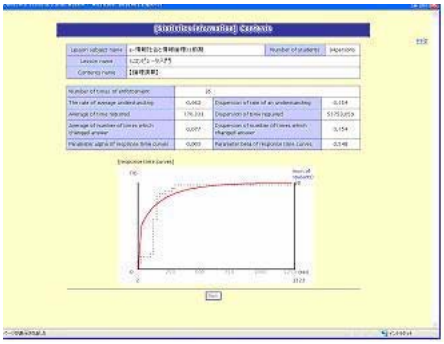

Figure 2.n line contents characteristics analysis system

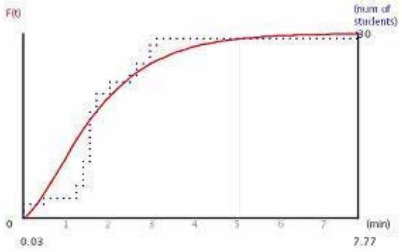

Figure 3. An example for test item content

(α =2.2, β =1.21)

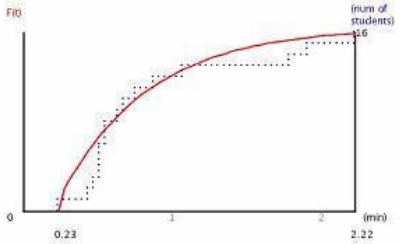

Figure 4. An example for animation content

α=0.81,β=1.42

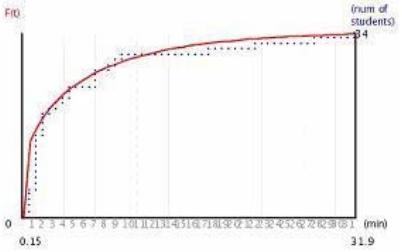

Figure 5. An example for Image content

(α =0. 428, = 0. 48)

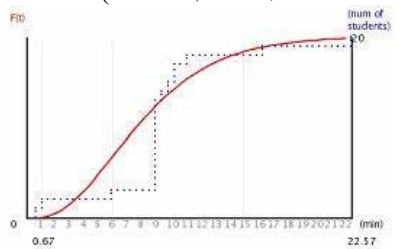

Figure 6. An example for text content

(α=3.595, β=0.5)

4.3. An example of $\alpha - \beta$ chart

The figure 7 is the $\alpha - \beta$ chart corresponding to 142 contents in the course "Computer".

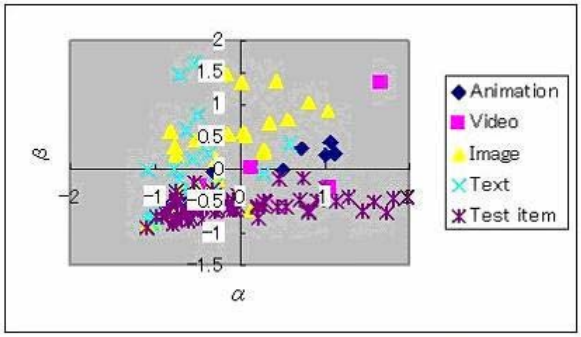

Figure7. An example of $\alpha - \beta$ chart

As the results, the following are derived:

- The animation contents in this course tend to be constructed by combining multiple simple contents which do not need long time.
- The video contents in the course are very simple contents which needs long time or the contents combining multiple simple contents which do not need long time.
- The image contents in this course tend to be constructed by combining many or few simple contents which do not need long time.
- The text contents in this course are very simple contents which need long time or not.
- The test contents in this course tends to be constructed by combining many or few simple contents which need long time.

Although this section demonstrate some examples of the response curves and $\alpha - \beta$ chart, it is the most important to improve the contents in e-learning by using this tool.

References

[1] M. Ueno, "Learning Log Database and Data Mining system for e-Learning", Proc. of International Conference on Advanced Learning Technologies 2002, *Proc. of ICALT2002*, pp.436-438, 2002

[2] M. Ueno, "LMS with irregular learning processes detection system", *Proc. of E-learn2003*, pp.2486-2493, 2003

[3] K. Nagaoka, A.Wu, "Analysis of Response Time data from Computer Testing", Japanese Journal of Educational Technology, Vol.12, No.4, pp.129-137,198