Learning Log Database and Data Mining system for e-Learning OnLine Statistical Outlier Detection of irregular learning processes

Article · January 2002

CITATIONS
9

READS
202

2 authors:

Maomi Ueno
The University of Electro-Communications
121 PUBLICATIONS **372** CITATIONS
SEE PROFILE

Keizo Nagaoka
Waseda University
28 PUBLICATIONS **93** CITATIONS
SEE PROFILE

Learning Log Database and Data Mining system for e-Learning -On-Line Statistical Outlier Detection of irregular learning processes-

Maomi Ueno
Nagaoka University of Technology
E-mail: ueno@kjs.nagaokaut.ac.jp
Keizo Nagaoka
National Institute of Multimedia Education
E-mail: nagaoka@nime.ac.jp

Abstract

Recently, distance education by using e-Learning has been popular in actual educational situations. However, there is a problem that the instruction strategy tends to be one way, and so it sometimes makes the learners bored comparing with usual instruction methods. This paper proposes a method of online outlier detection of learners' irregular learning processes by using the learners' response time data for the e-Learning contents. The unique features of this method are as follows: 1.It proposes an outlier detection method by using Bayesian predictive distribution. 2. It is available for small sample, 3.It is convenient to calculate the predictive distribution. 4.On-line learning is realized on WWW. 5.It assists two ways instruction by using data mining results for the learners' learning processes. The system was utilized for actual classes. The results show the efficiency of the system.

1. Introduction

Recently, distance education by using e-Learning has been popular in actual educational situations. However, there is a problem that the instruction strategy tends to be one way, and so it sometimes makes the learners bored comparing with usual instruction methods. To solve these problems, it is an important task to develop a new methodology of instruction based on the e-Leaning.

On the other hand, 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 save this data or how we utilize this data. Some researches has been proposed from this view point. Kawamura [1] and Matsumoto [2] proposed a clustering method for learners by using e-Learning log data. Matsui [3] proposed a data-mining method which constructs a tree by using the ID3 method. This paper proposes a method of online outlier detection of learners' irregular learning processes by using the learners' response time data for the e-Learning contents. The unique features of this method are as follows: 1.It proposes an outlier detection method by using Bayesian predictive distribution. 2. It is available for small sample, 3.It is convenient to calculate the predictive distribution. 4.On-line learning is realized on WWW. 5.It assists two ways instruction by using data mining results for the learners' learning processes. The system was utilized for actual classes. The results show the efficiency of the system.

2. e-Learning platform

The authors have developed an e-Learning platform system [4]. In this session, the outline of the system will be introduced. The platform system consists of 1. Contents Presentation System (CPS), 2. Contents Database (CD), 3.Learning Histories Database (LHD), and 4. Data Mining System (DMS). The CPS integrates various kinds of contents and present the integrated information on the web page shown in Figure 1. Moreover, the system presents some test items which confirm learners' comprehension degree as soon as the contents has been completed . An example of a test presentation is shown in Figure 2.

The CD is a database which consists of various kinds of medias, text, jpeg, mpeg, and so on. The proposed platform monitors learners' learning processes and saves them as a log data in the LHD. First, teacher makes the contents concerned with his lecture, and saves them in the CD. Then, the CPS automatically integrates the contents, and presents them to learners. The learners can learn them through the internet. The learners' learning histories log data is saved in the LHD, and it is analyzed in the DMS. The DMS presents the feedbacks for the learners and the teacher respectively. The teacher can know information about learners' learning processes, and he can give some comments or instructions to the learners by using e-mail.

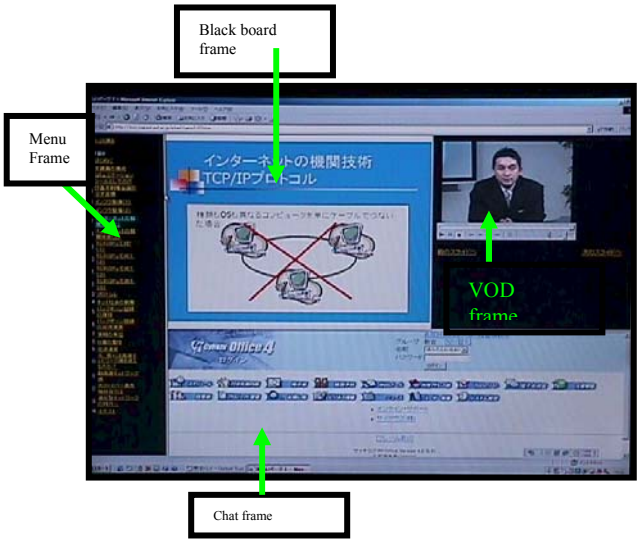

Figure 1. An example of e-Leaning platform presentation.

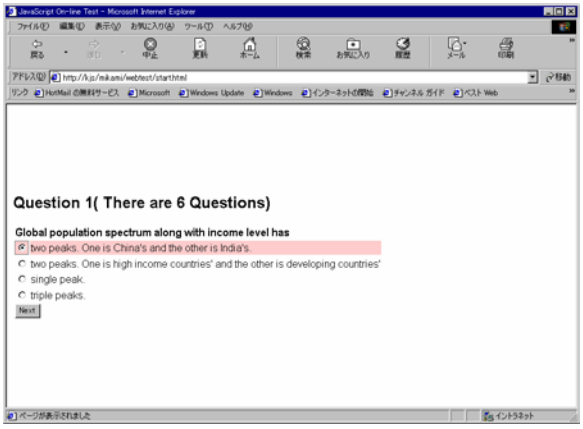

Figure 2. An example of Test frame.

3. Learning Log Database

The platform monitors learners' learning processes and saves them as a log data in the LHD. The saved data consists of A) Contents ID, B) Learner ID, C) The number which the learner has tried the content, D) Test Item ID, E) Operation order ID, F) Operation ID which indicates what operation was done in the content, G) Date and Time ID which indicates the time and date of starting the operation, and H) Time ID which indicates time that it is takes in the operation. The problem is how we efficiently use the huge amount of log data. The next section will propose one of utilization of this data, a method of online outlier detection of learners' irregular learning processes.

4. On-Line Statistical Outlier Detection

4.1. Model

The data which is used for the outlier detection is response time data. Here, a Bayesian predictive distribution of a new data x_{n+1} given the learner's learning processes data $x_1, \cdots, x_n$ will be derived for the outlier detection of learners' irregular learning processes. Let t_{ij} be a learner j's response time for the i-th content, and let consider the following linear equation

$$x_{ij} = \frac{t_{ij} - \bar{t}_i}{s_i} = \mu_j + e_j \quad , \qquad (1)$$

where

$\bar{t}_i$:The average of response time for the content i

$$\bar{t}_i = \frac{1}{m}\sum_{j=1}^{m} t_{ij} \qquad (2)$$

s_i :The standard division of of response time for the content i

$$s_i = \sqrt{\frac{1}{m}\sum_{j=1}^{m}(t_{ij} - \bar{t}_i)^2} \qquad (3)$$

μ_j :A parameter for the learner j

e_j : Random error

n: The number of the contents

m:The number of the learners

The error term follows the following the normal distribution.

$$e_j \leftarrow N(0, \sigma^2) \qquad (4)$$

Based upon the model (4), a Bayesian predictive distribution of a new data x_{n+1} given the learner's learning processes data $x_1, \cdots, x_n$ can be derived as follows;

$$p(x_{n+1} \mid X) = \int\int p(x_{n+1} \mid \mu, \sigma^2) p(\mu, \sigma^2 \mid x_1, \cdots, x_n) d\mu d\sigma^2 \qquad (5)$$

$$\propto \left(1 + \left[\frac{(x_{n+1} - \mu_*)}{\sqrt{\frac{n_0 + n + 1}{(n_0 + n)\nu}\lambda_*^{-1}}} \right]^2 / \nu \right)^{-\frac{\nu+1}{2}}$$

where

$$t = \frac{(x_{n+1} - \mu_*)}{\sqrt{\frac{(n_0 + n + 1)}{(n_0 + n)V}\lambda_*^{\,2}}} \qquad (6)$$

then t follows t distribution with degree of freedom $V = n_0 + n - 1$.

4.3 Outlier Detection

By using (6), we can detect outlier of learning processes.

The procedure is as follows:

1. Get a new data x_{ij}
2. Calculate the value of t in (6)
3. If t is greater than the value of t in t distribution with α or t is less than the value of minus t in t distribution with α, then the new data is detected.

If the data is detected, the teacher investigates the learner's learning processes, and sends e-mail with some comments.

5. Online outlier detection system

We actually used e-learning platform including this outlier detection system. The outlier detection system is shown in Figure 4. The system presents 1. learners names, 2. the learner's t value curve, and 3. the content with the irregular processes. We can know learners' learning processes on Online. We actually executed distance instruction by the e-learning platform including this outlier detection system. The figure 4 shows the learner 4's outlier detection curve. In this case, outlier processes appears the contents 129-145.The figure 5 shows the detection curve in the case that there are few outlier processes. The learner' processes for the content 82 were detected as too long response. The teacher discovered this process, and sent e-mail with the comment "Was the content 82 difficult? If you have any question, you can ask me anything."

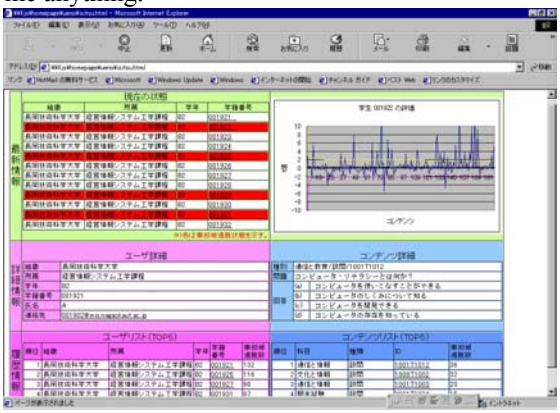

Figure 3. Online outlier detection system.

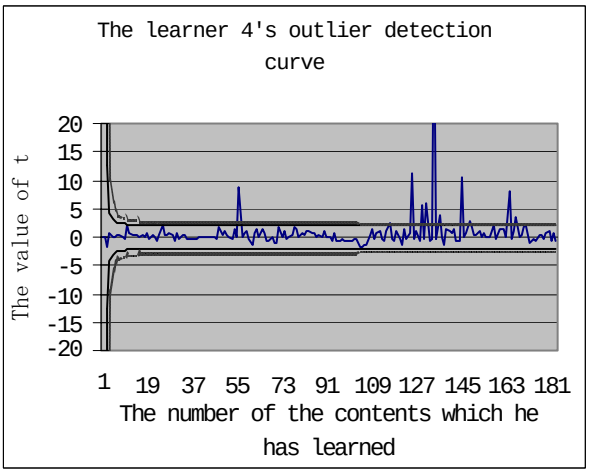

Figure 4. The learner 4's outlier detection curve.

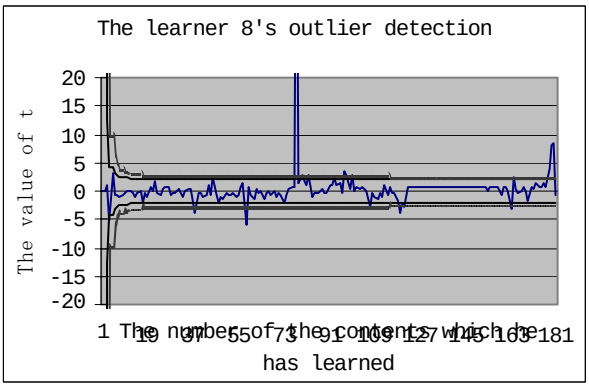

Figure 5. An example of detection curve in the case that there are few outlier processes.

6. References

[1] Kawamura, S., "A note on Learner model by using access Log to WWW server", ET, Vol 98,No,188, pp.17-24, 1998

[2] Matsumoto, J.," Learning histories analysis system for educational assistance", Journal of Information Processing, Vol.40,No.9,pp.3596-3607,1999

[3] Matsui, T and Okamoto, T.," Discovery Science of Digital Portforio, ET , ET2000-88, pp.87-94,2000

[4] Ueno, M, "e-Learning platform system based on learning history log database and data mining system", Proceeding of ICCE 2002. (submitting)