Correlation Analysis of Audio and Video Based on Large Data Mining Process of Massive Open Online Course

Xiaofei Xiao[1], Guang Chen[2], Huajie Sui[3, *]

[1]Higher Education Research Office, Jiangxi University of Traditional Chinese Medicine, Nanchang 330004, China

[2]School of Economics and Management, Jiangxi University of Traditional Chinese Medicine, Nanchang 330004, China

[3]School of Humanity, Jiangxi University of Traditional Chinese Medicine, Nanchang 330004, China

huajie_sui@yeah.net

*Corresponding author

Keywords: Data mining, Learning measurement, Target optimization, Relevance, Large data, MOOCs.

Abstract. In order to meet the demand of high quality scarcity learning resource and promote the learning autonomy and learning method innovation, it needs to analyze learning process data. Based on the relevance data mining methods, this paper designs the data mining system of learning measuring model, and this system can analyze large data in the learning process. Through data learning test for MOOCs large data, the variation curve of correlation accuracy rate is obtained, realizing the optimal allocation of learning methods and learning objectives, which provides a theoretical basis for the research on large-scale MOOCs online course.

Introduction

Massive Open Online Course (MOOCs) is a new emerging online learning method based on modern learning method, the internet and the mobile communication technology. According to the new definition of Gartner Company in 2012, large data has the following three characteristics: large quantity, fast, variety [1-3]. The large data generated by the MOOCs can carry out effective analysis of learning strategy model. This paper uses the relevance data mining model to do correlated analysis on data generated by MOOCs, which makes learning resources getting optimized configuration and provides a theoretical basis for the research on MOOCs.

Overview of Large Data Learning Strategy Model

Traditional learning measurement is the measurement of learning results, and the most representative is standardized test. But only the examination results can't understand the performance in the learning process, such as the learning speed, comprehend ability, interaction and cooperation with other learners etc. [4-6]. The tracking data of network course can better understand the learning behavior characteristics of learners, according to the measurement characteristic of large data, and it can evaluate the learning behaviors and results.
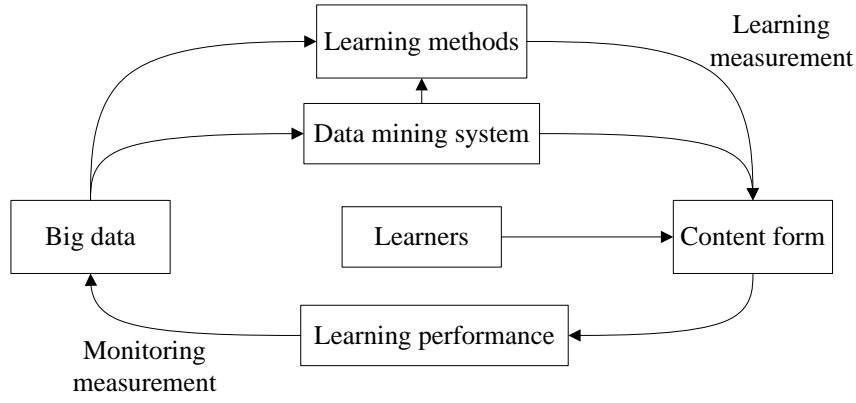
Fig.1: Large data learning measurement data mining model

Figure 1 shows the data mining mode of large data learning measurement model. It uses correlation analysis algorithm to design data mining system, based on the large data statistics of video content, image content and text content in the MOOCs learning, the best learning objective function is obtained.

The Design of MOOCs Data Mining Association

Learning objectives has a strong correlation with learning method, and the optimization of learning objectives needs to analyze the learning process, and the measured data is obtained. The online system based on MOOCs can carry large data measurement on learning process [7-9]. Assuming correlation function is $y = f(x_i)$, according to the correlation degree analysis of three times Hermite interpolation, $W_3(x)$ in the interval $[x_i, x_{i+1}]$ can be expressed as:

$$W(x_{i+\frac{1}{2}}) = \frac{1}{2}[f(x_i) + f(x_{i+1})] \qquad (1)$$

In order to realize the correlation data mining and get the optimization function of learning objectives, it needs to set the following boundary value:

$$\begin{cases} W_3(x_i - 0) = W_3(x_i + 0) \\ W_3'(x_i - 0) = W_3'(x_i + 0) \\ W_3''(x_i - 0) = W_3''(x_i + 0), \quad i = 1,2,...n-1 \end{cases} \qquad (2)$$

According to the three times spline interpolation model, the boundary value model can be obtained as shown in Figure 2:
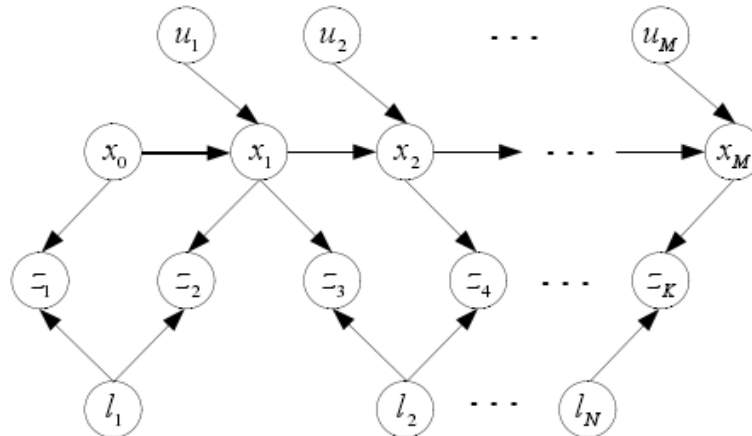

Fig.2: Boundary value conditions of association data mining

Figure 2 shows the boundary value conditions of association data mining. x represents learning measurement, l is the optimization goal, u is correlation function, then its endpoint needs to meet:

$$W_3(x_0 + 0) = W_3(x_n - 0), W_3'(x_0 + 0) = W_3'(x_n - 0)$$
$$W_3''(x_0 + 0) = W_3''(x_n - 0) \tag{3}$$

So at the two ends of $[x_i, x_{i+1}]$, there are:

$$W_3''(x_i) = \frac{6(y_{i+1} - y_i)}{h_i^2} - \frac{4m_i + 2m_{i+1}}{h_i} \tag{4}$$

It can be simplified to:

$$(1 - \lambda_i)m_{i-1} + 2m_i + \lambda_i m_{i+1} = \beta_i \tag{5}$$

According to the least square method, after linear, the optimization objective function of learning relevance measurement is shown in formula (6).

$$Q(i, j) = \sum_{k=1}^{N} [y_k - (i + jx_k)]^2 \tag{6}$$

Study on Large Data Learning Measurement Model based on MOOCs

In order to verify the effectiveness and reliability of learning measuring large data mining relevance algorithm designed in second section, this paper tests this method in the MOOCs learning measurement data analysis, and the design flow of learning measurement model is shown in Figure 3.
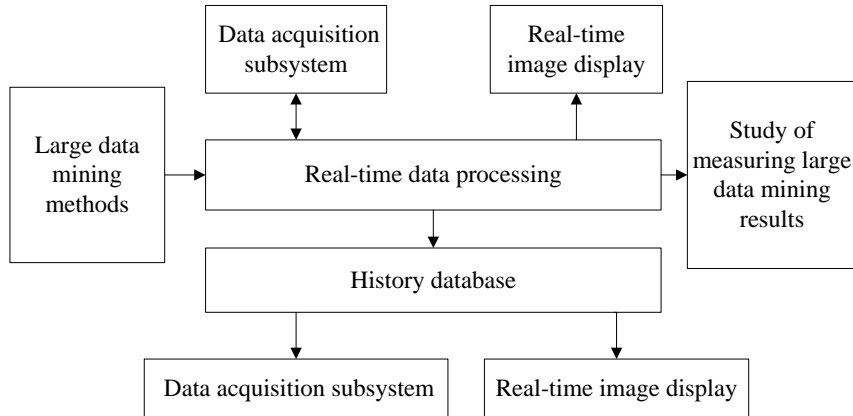

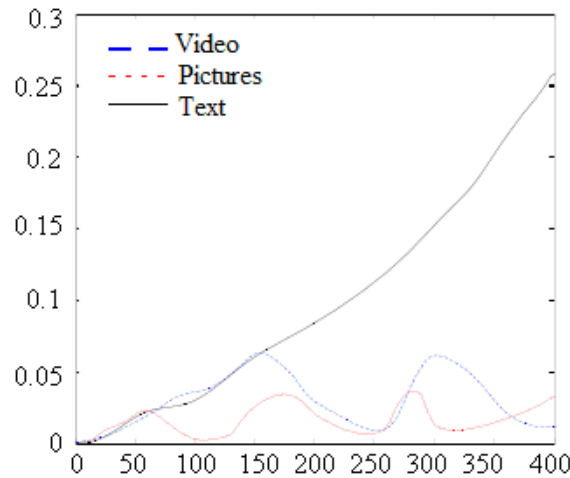Fig.3: Large data mining process for MOOCs system

Fig.4: Analysis of learning measurement data

Figure 3 shows the large data mining process for MOOCs system [10, 11]. The data mining process includes statistics of historical data and real-time data, and uses large data mining association analysis method to get the learned measurement data.

Figure 4 shows the MOOCs using large data mining on different learning content [12]. From the chart it can be seen, along with the increase of learning time, the use of pictures and text are less than the video significantly, and the use frequency has large fluctuation.
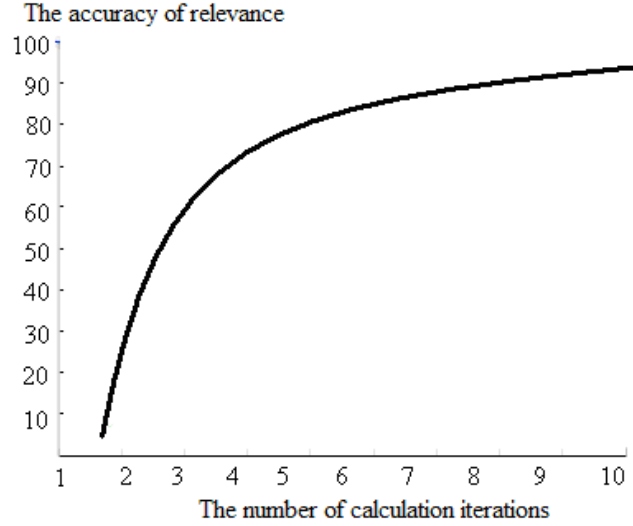

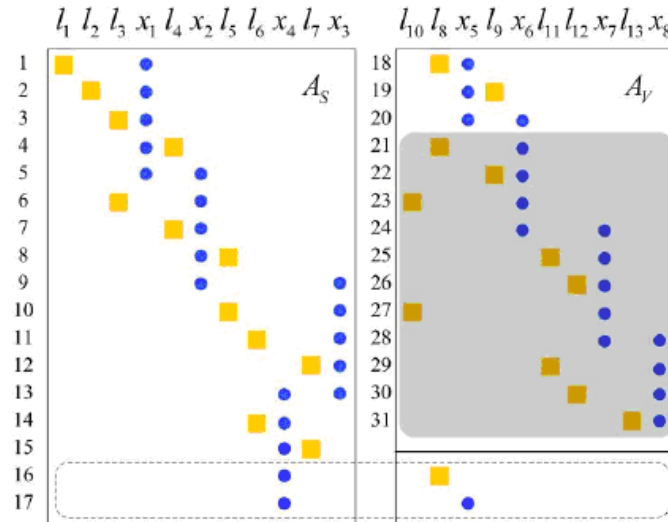Fig.5: The relevance accuracy rate statistical

Fig.6: Learning method relevance ranking

Figure 5 shows the calculation statistical curve of large data association accuracy rate [13]. From the chart it can be seen, when the computation iteration is up to 7 times, the large data association is more than 90%; when calculating iteration is up to 10, large data association is close to 100%, which reaches a high calculation precision.

Figure 6 shows the calculation results of large data relevance ranking for learning measurement model [14]. Through the data acquisition and analysis of MOOCs, and using MATLAB to do learning method distribution on 31 learning objectives, the learning method function is obtained, so as to achieve the learning objectives optimization purposes.

Summary

Based on data association mining methods, combined with three times spline interpolation boundary conditions, this paper designs the mathematical model of MOOCs large data analysis and

by setting the boundary condition, the optimized function of data mining model is obtained. Through the using frequency test of video, pictures and text in data mining system, the variation curve of correlation accuracy and the relevance ranking of learning methods are obtained, which realizes the optimal allocation of learning resources. But in terms of technology, how to realize the data sharing, achieve privacy protection and put the analysis in MOOCs system is the further work to be done in the future.

Acknowledgments

References

[1] W. Wang. Design and implementation of distance education platform based on web. China University of Geosciences, 2012: 2-16.

[2] D. Xie. Design and implementation of distance education platform based on web. Zhejiang University, 2012: 2-15.

[3] S.G. Poorman, C.A. Webb and M.L. Mastorovich. Students' stories how faculty helps and hinders students at risk. Nurse Educator, 2012, 2(27): 126-131.

[4] A.M. Gignac-Caille, M.H. Oermann. Student and faculty perceptions of effective clinicalinstructors in AD programs. The Journal of Nursing Education, 2013, 5(40): 347-353.

[5] X.F. Han. The present situation and the layering teaching thinking for teaching of higher mathematics. Journal of Lvliang College, 2012, 23(1): 40-41.

[6] L. Fu. Application of hierarchical teaching in English teaching. Hope monthly, 2012, 3(8):100-101.

[7] X.Q. Yu. Causes and breakthrough of Matthew effect in the basic education. Teaching and management, 2012, 5(9): 60-61.

[8] X.Q. Wang. The hierarchical teaching theory and practice of higher mathematics model. Education in Jilin, 2013, 3(11): 20-21.

[9] D.H. Mao. Implementation of the hierarchical teaching in higher vocational mathematics. Journal of Chongqing Vocational and Technical College, 2012, 16(4): 43-44.

[10] B.P. Chen, A.L. Hu. Hierarchical teaching mode in higher education massification stage. Liaoning radio and Television University, 2012, 3(2): 51-52.

[11] F.Y. Li. The application of multimedia in English teaching. Chinese information technology, 2012, 2(8): 43-45.

[12] L.J. Song. Exploration of the psychological characteristics in English teaching. Education exploration, 2013, 4(3): 55-56.

[13] P. Zhang, H.Y. Zhao. The network learning monitoring research. Chinese education information, 2012, 3(15): 56-59.

[14] J.F. Xu. Research on autonomous learning in thirty years: review and reflection. The seventh national language forum, 2012.