Stage Predicting Student Stay Time Length on Webpage of Online Course Based on Grey Models

Qingsheng Zhang, Kinshuk, Sabine Graf, Ting-wen Chang[1]

Faculty of Science and Technology, Athabasca Univeristy, Canada
[1]National Chung Cheng University, Taiwan
{qzhang,kinshuk,sabineg}@athabascau.ca, [1]ctw@cs.ccu.edu.tw

Abstract. To provide adaptive learning, an e-learning system needs to gather information about what student state is while the student learns online course. A student state index is the length of time the student stays on a webpage of online course. By predicting student's stay time length, the e-learning system has potential to dynamically tailor the learning content to the students. A literature review is conducted on power law of learning and knowledge component. We assume that online course consists of knowledge components. A knowledge component crosses several successive web pages. Accordingly, an initial prediction method is proposed based on the two learning curve modes and the grey models. Based on the experimental result of this initial prediction method, construction method of grey models is modified. The results produced by the grey models based on the two construction methods are then compared and analyzed. The results show that prediction of stay time length is possible to certain degree while the students learn knowledge on web pages.

Keywords: Stage Prediction, Learning Curve Mode, Knowledge Component, Power Law, Grey Model

1 Introduction

In e-learning, students' various characteristics, such as learning style, as proposed by [1], can be modeled for adaptive learning. In this paper, stay time length is considered for adaptive learning. Stay time length is the time period that a student stays on a webpage of online course. It is affected by instructional organization of knowledge on web pages, the knowledge level and the cognitive ability of student. We explore if stay time length can be predicted based on the most recent history data of students' stay time length. However, in the context of e-learning, stay time length does not always mean that students are actually learning while they are on a webpage. Such duration is processed as noise. If the proposed method is able to predict stay time length, detailed research can be conducted in future about filtering noise.

The phenomena of power law are pervasive in the physical world and social environments. One of the earlier observations of the phenomena of power law in learning process is by [2]. Power law of learning states that to learn a skill, the learning performance and learning opportunities follow the distribution of power law. The power law of learning ranges from short term perceptual tasks to team-based

longer term tasks [3]. Studies have also shown that the power law relationship between practice and performance may not always be readily evident in the practice of complex skills [4], but by decomposing the skill required in practicing each of the underlying rules, a cl ear power law relationship can be seen between practice and performance. If complex knowledge is decomposed into basic units, the fine grained analysis shows that learning curves of these basic units follow power law [5]. This kind of basic unit is named as knowledge component (KC). KC refers to the abstract knowledge unit containing concepts, principles, rules, declarative knowledge and schemata. In this paper, it is assumed that online course consists of KCs, and a KC crosses several successive web pages.

The rest of paper is organized as follows. The learning curve modes, power law and two grey models are introduced in section 2 i n order to understand how to forecast stay time length. Lastly, the paper is concluded with a summary of the findings and discussion on future direction.

2 Learning Curve Modes, Power Law and Grey Models

A power law can be viewed as a probability distribution function in which the production of probability that a random variable takes a value is proportional to a negative power of this value. It is described by formula 1:

$$P(x) = c * x^{-k} \qquad c > 0, k > 0 \qquad (1)$$

In the context of power law of learning, p is performance in formula 1, x is practice opportunities, c is performance on the first opportunity, and $-k$ is learning rate. With an increase in the practice opportunities, the performance decreases greatly at the beginning, and after some opportunities, the performance arrives at a s table state. Based on the assumption in section 1, we investigate if power law of learning exists in the log data of online course. Fig. 1 shows the stay time length curves of the learning of two students over two days. Curves in Fig. 1 show that fluctuations in stay time length varied greatly.
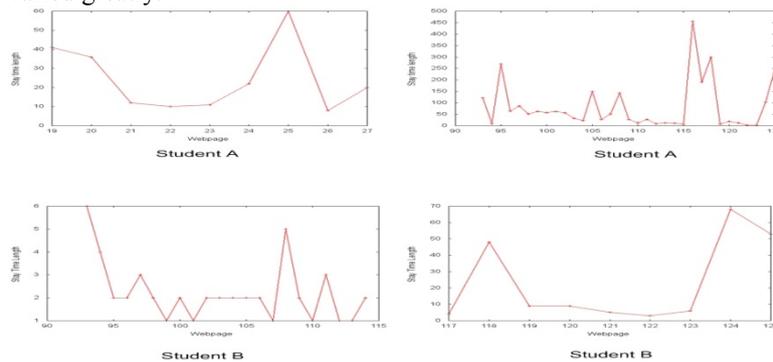

Fig. 1. Stay time length curves of two students learning in two days

In Fig. 1, y axis is stay time length on a webpage, and x axis is the index of web pages. Each diagram in Fig. 1 displays the record of stay time length in one day for the students A and B. The overall trend of stay time length in each curve is totally

different from power law curve. But some parts of each curve in Fig. 1 are similar to power law curve to some degree. We also observe that other parts of each curve in Fig. 1 cannot be described by formula 1. However, the stay time length of these parts roughly follows another power law while the index k in formula 1 becomes negative. Therefore, we proposed two possible learning curve modes and their variants (broken line) as shown in Fig. 2, where k is webpage index, and y axis is stay time length.
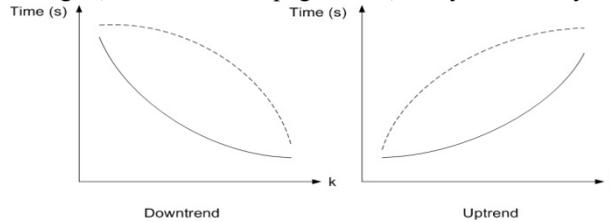
Fig. 2. Two learning curve modes and their variants

As proposed by [2], power law may be an artifact arising from averaging and that the exponential function may be the best fit when individual subjects employing a single strategy are considered. We substitute formula 2 for formula 1.

$$T(x) = c * e^{-mx} \qquad c > 0, m > 0 \qquad \textbf{(2)}$$

The trend of power law in the parts of curve in Fig. 1 is not evident. To predict stay time length, the grey models [6] is used. Advantage of using grey models is that they only require a few of history data for prediction. The accumulated generating operation (AGO) of grey models can make the unobvious trend of power law in the parts of curve in Fig. 1 evident. The AGO is defined as follows. Let $L^{(0)}$ be the original non negative sequence and $l^{(0)}$ be an element in the sequence. The AGO sequence $L^{(1)}$ can be presented as follows:

$$l^{(1)}(k) = \sum_{i=1}^{k} l^{(0)}(i) \qquad k = 1, 2, \ldots, n \qquad \textbf{(3)}$$

where $l^{(1)}$ is an element in the sequence $L^{(1)}$. To bring the data back to the original state, the inverse AGO (IAGO) is applied to the current data sequence. AGO and IAGO are a pair of inverse sequence operators. IAGO operation is defined as follows:

$$l^{(0)}(1) = l^{(1)}(1)$$
$$l^{(0)}(k+1) = l^{(1)}(k+1) - l^{(1)}(k) \qquad k = 1, 2, \ldots, n \qquad \textbf{(4)}$$

The curves used to fit AGO sequences can be represented by differential equations. These differential equations are referred as grey models, such as GM (1, 1) and Verhulst. The GM (1, 1) is often used to fit exponential type sequences. Verhulst is used to fit sequences with saturated trend. The typical curves of the two models are illustrated in Fig. 3, where k denotes the webpage and $l^{(0)}(k)$ denotes stay time length.
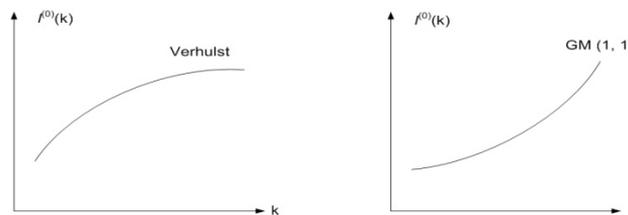
Fig. 3. Curves of model GM (1, 1) and Verhulst

Only using GM (1, 1) may not appropriately predict the stay time length. Therefore, we also use Verhulst model to predict stay time length. Now, we see that two grey models can only fit the learning curve mode of right side in Fig. 2. For the learning curve mode of left side in Fig. 2, we use the technique of curve flip to convert original curve into the curve which can be used by the two grey models. GM (1, 1) means grey model with single variable in first order equation. A type of single variable and first order linear dynamic differential equation is shown as follows:

$$\frac{dL^{(1)}}{dt} + a * L^{(1)} = b \qquad (5)$$

where $L^{(1)}$ is the AGO sequence of the original sequence, and the parameters a and b are referred as the development coefficient and grey action quantity respectively, which can be acquired as follows:

$$\hat{a} = \begin{bmatrix} a \\ b \end{bmatrix} = (B^T B)^{-1} B^T Y_N \qquad (6)$$

where

$$B = \begin{bmatrix} -z^{(1)}(2) & 1 \\ -z^{(1)}(3) & 1 \\ ... & ... \\ -z^{(1)}(n) & 1 \end{bmatrix} \qquad (7)$$

$$Y_N = \begin{bmatrix} l^{(0)}(2) \\ l^{(0)}(3) \\ ... \\ l^{(0)}(n) \end{bmatrix} \qquad (8)$$

and

$$z^{(1)}(k) = \frac{l^{(1)}(k) + l^{(1)}(k-1)}{2} \qquad k = 2, 3, ..., n \qquad (9)$$

Then, the time response sequence of GM (1, 1) can be written as follows:

$$\hat{l}^{(1)}(k+1) = (l^{(0)}(1) - \frac{b}{a}) e^{-ak} + \frac{b}{a} \qquad k = 1, 2, ..., n \qquad (10)$$

where $\hat{l}$ is the predicated l. $\hat{l}^{(0)}(k+1)$ can be acquired by IAGO as follows:

$$\hat{l}^{(0)}(k+1) = \hat{l}^{(1)}(k+1) - \hat{l}^{(1)}(k) \qquad k = 1, 2, ..., n \qquad (11)$$

The Verhulst model is a t ype of single variable and second order differential equation, as follows:

$$\frac{dL^{(1)}}{dt} + a * L^{(1)} = b * (L^{(1)})^2 \qquad (12)$$

where a and b are also acquired with formula 6. Y_N in formula 6 for Verhulst model is the same to formula 8, and B in formula 6 can acquired as follows:

$$B = \begin{bmatrix} -z^{(1)}(2) & (z^{(1)}(2))^2 \\ -z^{(1)}(3) & (z^{(1)}(3))^2 \\ ... & ... \\ -z^{(1)}(n) & (z^{(1)}(n))^2 \end{bmatrix} \qquad (13)$$

$z^{(1)}$ in formula 13 is the same to formula 9. For Verhulst model, its time response sequence can be presented as follows:

$$\hat{l}^{(1)}(k+1) = \frac{a * l^{(1)}(0)}{b * l^{(1)}(0) + (a - b * l^{(1)}(0))e^{ak}}$$

$$k = 1, 2, ..., n \qquad (14)$$

As saturated sequence already has some degree of exponential form, so the original sequence can be regarded as $L^{(1)}$ directly. $L^{(0)}$ in equation (8) should be the IAGO sequence of the original sequence. For Verhulst model, the predicted values of the original sequence can be acquired from equation (14). The two grey models are used to predict stay time length based on most recent three history data items. First, an extracted subsequence in raw sequence is evaluated for its curve type, and chooses a grey model for prediction. Then, we shift the subsequence window in the sequence to next data for new subsequence. A chosen grey model is used for next step prediction. This process is repeated until the last subsequence. As the abovementioned, we call our prediction method as stage prediction.

3 Experiment Analysis and Conclusion

Based on the prediction models in section 2, only data from the instructional unit lesson of online course are used, with learning events 91,084 from 459 s tudents. Noisy low value (NLV) is set to less than or equal to 2 seconds. Noisy high value (NHV) is set to larger than or equal to 300, 1200 and 5400 seconds respectively. More than 50,000 learning events are used for testing while noise is filtered. For each prediction, we compare the predicted value with the original value to obtain the evaluation result. Here, we use the mean magnitude of relative error for all students (AMMRE) as evaluation measure. By averaging magnitude of relative error over number of predictions n in all days per student and all students m, AMMRE is obtained from the following formula:

$$\text{AMMRE} = \frac{1}{m}\sum_{j=1}^{m}\left(\frac{1}{n}\sum_{i=1}^{n}\left|\frac{p_i - \hat{p}_i}{p_i}\right|\right)_j \qquad (15)$$

where $\hat{p}_i$ is the predicted value of original value p_i.

Table 1 shows the AMMRE in different NHVs. The higher value in the column AMMRE means the lower prediction accuracy. NHV takes serious effect for prediction. We also inspect prediction results in detail as we use the method of one-step-subsequence to construct the grey models. The underlying assumption is that it can dynamically fit the change trend of stay time length.

Table 1. AMMRE of prediction for different NHVs.

NLV(s)	NHV(s)	AMMRE (%)	Predicted Learning Events
2	300	459.89	51,055
2	1200	997.37	57,060
2	5400	2169.89	59,231

One KC crosses several successive web pages. If a student does not know this KC, the student spends more time at the beginning. The stay time length would decrease for the subsequent web pages once the student has acquired the basic understanding. After the student finishes this KC on these web pages, the subsequent web pages may start another KC. It can drastically change the trend of stay time length. Therefore, the prediction is significantly affected. It reveals that the method of one-step-subsequence

cannot dynamically reflect the occurrence of a n ew KC. Therefore, the proposed modified method observes the ratio between the predicted value and the original value. If the ratio exceeds the predefined value, the prediction models are built with stay time length from new KC. For the modified method, AMMRE changes from 37.45% to 121.06% in different NHVs. The value of AMMRE mainly depends on the ratio. The predicted learning events are 105, 4355, 5622, 5961, 5916, 5828 and 5706 as the ratio changes from 1 to 7 while NHV is 5400. As three most recent history data are used for prediction, in the worst situation, about 10 percent predicted learning events in Table 1 are used. In the best situation, we construct model for each predicted learning event, which means that about 30 percent predicted learning events in Table 1 are used besides ratio 1.

In this paper, we propose a stage prediction method about stay time length according to the learning curve modes and grey models. The initial prediction method is tested. Then, based on the experimental analysis, the analysis shows that the construction method of grey models is not robust to fit the change trend while a new knowledge component occurs. Therefore, we modify the construction method. The results show that AMMRE does not have strong effect from NHVs. It is dependent on the predefined ratio. In this paper, we consider only stay time length. Other factors, such as the length of webpage, may be used in future to model the stay time length. Additionally, we explore whether alternative learning curve modes exist in the dataset.

Acknowledgement

References

1. Peter, B.: Adaptive Hypermedia. User Modeling and User-Adapted Interaction. 11(1/2), pp 87--110. (2001)
2. Seibel, R.: Discrimination reaction time for a 1,023 alternative task. Journal of Experimental Psychology. 66(3), pp 215--226. (1963)
3. Ritter, F.E., Schooler, L. J.: The learning curve. In: W. Kintch, N. Smelser, P. Baltes, (Eds.), International Encyclopedia of the Social and Behavioral Sciences. Oxford, UK: Pergamon. pp 8602--8605. (2001)
4. Kenneth, R. K., Santosh, M.: Distinguishing Qualitatively Different Kinds of Learning Using Log Files and Learning Curves. In: ITS 2004 Log Analysis Workshop, Maceio, Brazil. pp 39--46 (2004)
5. Hausmann, R., VanLehn, K.: Self-explaining in the classroom: Learning curve evidence. In McNamara & Trafton (Eds.), In: the 29th Annual Cognitive Science Society, Austin TX: Cognitive Science Society, pp 1067--1072. (2007)
6. Sifeng, L.: The Grey System Theory and Applications (in Chinese). Beijing, China: Science Press. (2005)