

Evaluation of Student Performance with Predicted Learning Curve Based on Grey Models for Personalized Tutoring

<https://doi.org/10.3991/ijet.v14i13.9880>

Manqiang Liu

Lanzhou University of Technology, Lanzhou, China

Qingsheng Zhang ^(✉)

Xi'an University of Posts and Telecommunications, Xi'an, China

zhangqs@xupt.edu.cn

Abstract—Learning time of student is precious, over-practice of target knowledge component wastes student's time, however, under-practice may mean the student may not grasp target knowledge component properly. To any student, it is helpful if intelligent tutoring system can determine how many practice opportunities needed for mastery of knowledge component. In this paper, to improve student's learning efficiency, a method of predicted learning curve based on grey models is proposed to determine the counts of practice opportunity for mastery of knowledge component. The experimental results show that the predicted value on error rate of practice opportunity over knowledge component with the proposed method is much closer to the value of real learning curve than the predicted learning curve produced by learning factors analysis. It implies the proposed prediction method is potential to present reasonable practices for personalized tutoring.

Keywords—Knowledge component, power law, predicted learning curve, learning factor analysis, grey models, personalized tutoring

1 Introduction

Advances in computerized educational technologies, students and teachers like to facilitate their learning and teaching with information technologies in the instructional process. For example, learning analytics is used to study how closely student following the suggested sequence of courses in a degree curricula and student's performance as stated in [1]. Another example of educational technologies as stated in [2] shows that the statistical analysis and visualization based on techniques of association rule and clustering can help teachers to understand how students learn and improve their learning online learning process. The quantitative relationship between students' browsing behavior and the web pages is mined with a multivariate regression, potentially interested content can be recommended for students as stated in [3]. A learning path constructing method is available by using formal concept analysis to generate

concept lattice with the keywords extracted from student's learning resources, then an optimized learning path based on the learning objects recommended with preference and correlation-based algorithms is presented for students as stated in [4]. Also, many studies about learning spaces technology as stated in [5] are surveyed for what learning space can support dynamic, engaged and inspired learning behaviors of students.

In a variety of computerized educational technologies, one of most popular and successful candidates is intelligent tutoring system (ITS) as stated in [6]. ITS can be used to improve the student's performance with practice opportunities. Dynamic Bayesian networks employed by ITS application as stated in [7] can evaluate students' algebraic knowledge with the probabilities based on the defined configuration parameters of student's initial knowledge level, guess rate, slip rate and learning rate and decide to do what next practice. In another ITS application as stated in [8], cognitive efficiency of student is assessed by the difference between the standard score of performance and mental effort rating of student, adaptive presentation strategy based on the cognitive efficiency releases alone or combined practices of database structured query language to students for fewer learning activities.

Generally, ITS tracks the knowledge of student, and evaluates what degree of knowledge component (KC, that is a skill, concept, or a fragment of information that is used to carry out one task, often along with other KCs) is mastered by the student with offering practice opportunities and determines what KC in next step as stated in [9]. To better student's learning efficiency, only necessary and minimum practices of the same KC should be provided if the performance of student proves that the student mastered the KC as stated in [10]. Predicted error rate in next practice opportunity of the same KC is lower than 20%, which was considered as mastery threshold, this KC is basically mastered by the student, and thus a new KC will be presented to the student by tutor as stated in [11]. Therefore, a key issue in ITS application is how to predict student's performance for next step practice of KC. In this paper, a predicted learning curve produced by a kind of machine learning method named grey model is used to predict student's practice performance of KC for the learning efficiency of student.

The rest of paper is organized as follows. In section 2, literature review about learning curve is presented. In section 3, one of the additive factor models and grey models for predicted learning curve are separately discussed. The experiment and discussion are analyzed in section 4. The paper is concluded with a summary of the findings in section 5.

2 Literature Review

The interactive association between the error rate in performance and opportunities of practice is delineated by power law as stated in [12]. Generally, the association exhibits that the error rate may decrease along with the amount of practice increasing. This function between the error rate and amount of practice is known for "learning curve". In fact, the phenomenon of power law is quite pervasive in the world, such as the relationship between number of cities and population in a country. Power law also

occurs in modules comprising software ranging from simple java classes to systems using self-contained libraries written in C, Perl and Ruby as stated in [13], that is, the numbers of modules (i.e. class or function) and their references follow power law. One of the earlier observations was reported for the phenomena of power law even occurring in learning process as stated in [14]. Power law in learning states that to acquire a skill, the learning performance and practice opportunities generally follow the distribution of power law. Power law in learning is omnipresent, ranging from short term tasks about human perceptual to group-based longer term tasks of human social behavior in breadth and length, the error rate that people experience decreases with practice opportunities pursues a similar pattern as stated in [15]. However, some studies have also presented that this kind of power law association between practice and performance is not always readily distinct for those practices of complex skills as stated in [16], but the decomposition of skill required in practicing each underlying KC is applied, a clear power law curve is seen as function between practice and performance.

In practice, a learning curve presents variations in student performance along with practice opportunities. The real learning curves (that is, average observed errors of a KC over each practice opportunity) plotted with the data containing much noise may present the shape of undulating lines. The learning curve predicted with additive factor model as stated in [17] is much smoother. A set of customized models based on item response theory are used in this additive factor model to predict what performance a student may have for each KC on each practice opportunity. The predicted learning curves are plotted with the average predicted error of a KC over each of the learning opportunities. Due to filtering much of the noise in the log data on student learning with the additive factor models, the predicted learning curves appear much smoother than the real learning curves. The real learning curve may give a visual clue about how well a student may perform over a set of practicing opportunities, the predicted curves allow for a more precise prediction of a success rate at any learning opportunity. The predicted learning curve can be used to measure how much practice is needed to master a KC for efficiency, that is, the student is not over-practice or under-practice for mastery of a KC. However, this prediction method based on additive factor model is only used in statistical style for all students. For example, KCs in curriculum can be designed again based on this kind of predicted learning curve. Therefore, the predicted learning curve based on additive factor model is not effective for individual student personalized tutoring. In this paper, an alternative method for predicted learning curve based on grey models is started to explore personalized student tutoring.

3 Methodology

3.1 Power law curve

Power law is considered as a probability distribution function, that is, the production of probability that a random variable takes a value is proportional to a negative power of this value. It is presented in Equation (1):

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

While power law is linked to learning curve, the parameters Equation (1) are explained as follows: $P(x)$ is error rate, x is practice opportunities, c is the error rate on the first attempt of target KC's practice opportunities, reflecting the difficulty level of target KC; k is the learning rate, reflecting how easy a KC is learned by student.

A smooth curve in Figure 1 is plotted according to a power function with given parameters in Equation (1). The other in Figure 1 is plotted based on student's real performance data. The error rate decreases greatly at the beginning after several practice opportunities, and the performance arrives at a stable state with lower error rate after more opportunities. Power law is typically applied to individual KCs. As the cognitive model in ITS application uses many KCs and has the data about many students, the power law model is extended in this context [18], which is called Learning Factors Analysis (LFA), one of additive factor models. The information from LFA may be used to improve the tutor and the curriculum, for example, instructor can identify over-practiced or under-practiced KCs, or individual student over-practiced for specific KC.

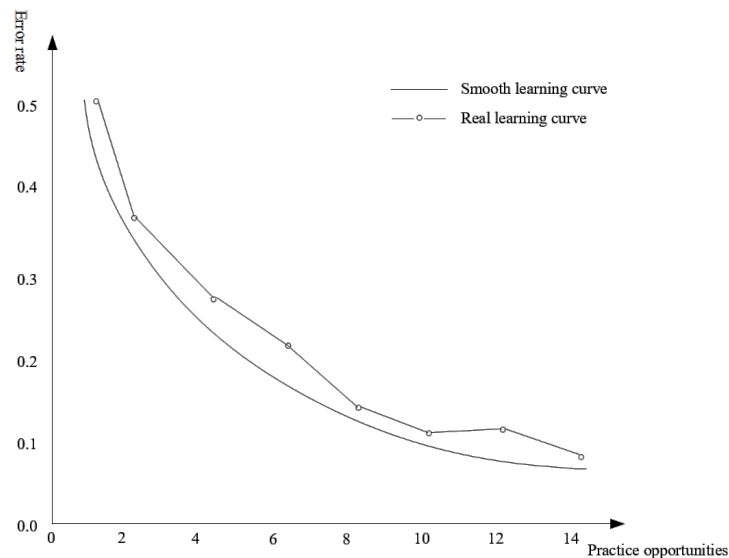


Fig. 1. Power law and learning curve

3.2 Learning factors analysis for predicted learning curve

To quantifiable analysis of student learning KC, a multiple logistic regression model is developed, in this LFA model, the dependent variable is whether the student answers the question of KC correctly or not. Since a different initial error rate on a KC for each student is estimated, student as one of the predictor variables is accounted for the dependence. The other predictor variables are KCs, and the interaction between KCs and students. This regression in Equation (2) is shown as follows:

$$\ln \left[\frac{p_{ij}}{1-p_{ij}} \right] = \sum \alpha_i * X_i + \sum \beta_j * Y_j + \sum \gamma_j * Y_j * T_j \quad (2)$$

p_{ij} - the probability that student i knows a KC_j correct,

α_i - the coefficient for each student i ,

X_i - the covariates for each student i , $\alpha_i * X_i$ is the overall proficiency of student i ,

β_j - the coefficient for each KC_j , that is, difficulty factor of KC_j ,

Y_j - the covariates for KC_j , 1 for being present, 0 for being absent, $\beta_j * Y_j$ is the difficulty level of KC_j ,

γ_j - the coefficient for the interaction between a KC and its practice opportunities,

T_j - the covariates for the counts of opportunities practiced on KC_j , $\gamma_j * Y_j * T_j$ is the mount gained for each opportunity.

In Figure 2, it shows the real learning curves in red based on the log data of students and the predicted learning curves in blue based on Equation (2). In Figure 2-1, Learning curves are plotted for KC model “Area”, where data about all related sub-KCs and all students are used. In Figure 2-2, Learning curves are plotted with data of one sub-KC “Area formula” from KC model “Area” and all students. In Figure 2-3, learning curves are plotted with data from KC model “Area” and one student. In Figure 2-4, Learning curves are plotted for KC model “Original”, where data about all related sub-KCs and all students are used. In Figure 2-5, Learning curves are plotted with data of one sub-KC “Trapezoid area” from model “Original” and all students. In Figure 2-6, Learning curves are plotted with data from KC model “Original” and another student. All charts in Figure 2 are plotted with data from DataShop [19].

In Figure 2, with simple visual inspection, it concludes that the prediction performance of LFA model as a statistical method is better for those data from more sub-KCs and students than for data from less sub-KCs and students. Especially, individual student learning curves greatly differ from the predicted learning curve. Therefore information from LFA can generally be used to improve the design of KCs in curriculum and tutor. To individual student’s personalized learning, a new method for prediction of error rate is needed. Therefore, two grey models are introduced in this paper to fit real learning curve of student for possible personalized tutoring application in ITS.

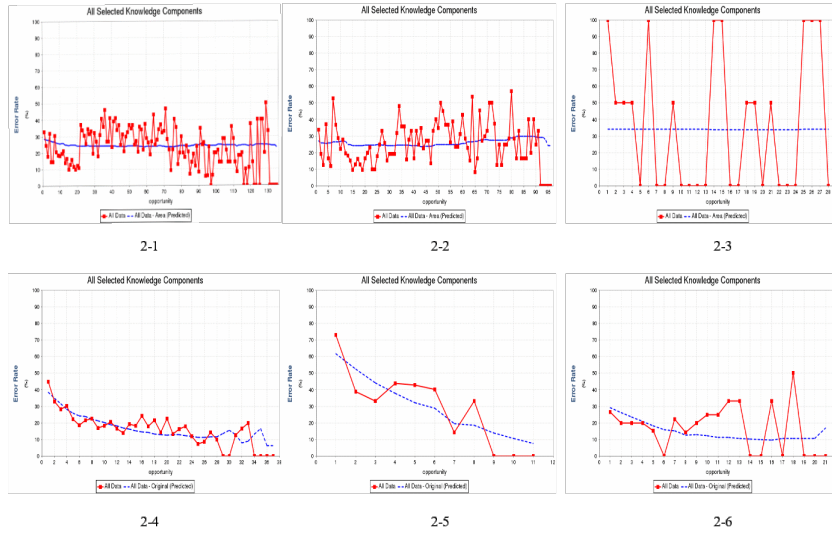


Fig. 2. Real learning curves in red and predicated learning curves based on LFA in blue

3.3 Grey models for predicted learning curve

Segments in learning curves in Figure 2 present uptrend or downtrend, and two types of trends are switched between them now and again. Therefore two learning curve modes in Figure 3 are developed for uptrend and downtrend segments in learning curve, and they are applied to the predicting error rate on practice opportunities of student with the grey models. The learning curve mode in Figure 3-1 shows that the error rate of each opportunity in student exploring target KC decreases from high to low. Another learning curve mode in Figure 3-2 shows that the error rate of each opportunity in student learning target KC rises from low to high.

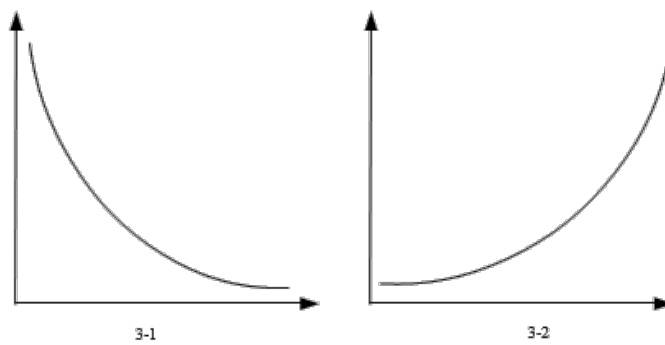


Fig. 3. Two learning curve modes

Power law is an artifact based on averaging, and the exponential function may be the best fit for power law as stated in [14]. Therefore, Equation (3) substitutes for Equation (1) for the simplicity of predicting error rate.

$$T(x) = c * e^{-mx} \quad c > 0, m > 0 \tag{3}$$

Some line segments in each real learning curve in Figure 2 are roughly similar to two learning modes in Figure 3. The trends of line segments in Figure 2 being similar to the patterns of learning curve modes are not even evident. To improve the performance of predicting error rate and more simplicity, the grey models of grey system theory is used instead of Equation (3). Grey models only require a few items of historic data for prediction in next step, and the accumulative operation of grey models can make the unobvious trends being similar to two learning curve modes in Figure 2 evident. The grey models are widely applied to the prediction in a variety of natural and social issues, such as transportation effects evaluation on air quality trends as stated in [20]. Generally, the nonnegative raw data sequence has no obvious patterns in uncertain circumstance as stated in [21], and it is difficult to find a proper curve or function for its fitting, however, applying an accumulated generating operation (AGO) to such raw data sequence, the resultant sequence monotonically increases and has a strong exponential characteristic, then an optimum exponential curve can be found to fit the trend of original sequence.

The following is the definition of AGO, let $L^{(0)}$ be the original non negative sequence, $l^{(0)}$ be an element in the sequence, then AGO sequence $L^{(1)}$ can be presented as follows:

$$l^{(1)}(k) = \sum_{i=1}^k l^{(0)}(i) \quad k = 1, 2, \dots, n \tag{4}$$

In Equation (4), $l^{(1)}$ is an element in the sequence $L^{(1)}$. If the produced exponential characteristic in Equation (4) is not yet apparent, AGO operations is applied to the sequence again. To recover the original state of data from the accumulative state, the operation of inverse AGO (IAGO) is applied to the current data sequence. AGO and IAGO are a pair of inverse sequence operators. IAGO operation is defined in Equation (5) as follows:

$$l^{(0)}(1) = l^{(1)}(1)$$

$$l^{(0)}(k + 1) = l^{(1)}(k + 1) - l^{(1)}(k) \quad k = 1, 2, \dots, n \tag{5}$$

In the grey system theory, the curves fitted by AGO sequences can be represented with differential equations. These differential equations are referred to as grey models. A variety of grey models corresponding to the different kinds of AGO sequences are used for the prediction in application fields. One of them is the grey model GM (1, 1) for fitting data sequences with exponential type, which describes any monotonic increasing procedure. Another grey model Verhulst is used for fitting sequences with saturated trend. The typical curves of two grey models are plotted in Figure 4, where k denotes the practice opportunities and $l^{(0)}(k)$ denotes the error rate in the context of predicted learning curve.

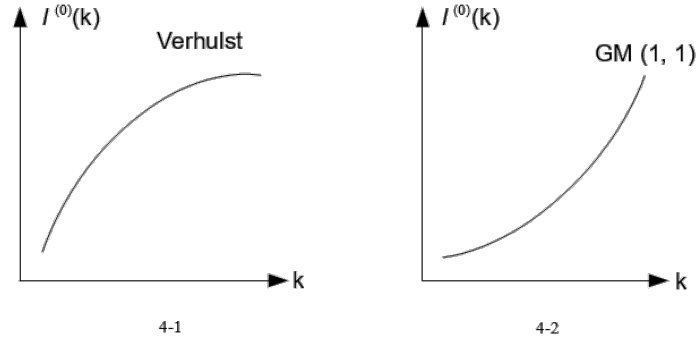


Fig. 4. Typical curves for grey models GM (1, 1) and Verhulst

Due to the complexity of student learning process, merely depending on GM (1, 1) is not appropriate for predicting the error rate, therefore Verhulst Model in Figure 4-1 is also used to predict the error rate. Now the two grey models only have capacity to fit the curves being similar to the learning curve mode in Figure 3-2. For the curves being similar to learning mode in Figure 3-1, the technique of curve flip is used to convert original curve into the curve which can be used by the two grey models. The detail of this procedure is described in the subsection 3.4.

GM (1, 1) states a grey model with single variable in first order equation. A type of single variable and first order linear dynamic differential equation is shown in Equation (6) as follows:

$$\frac{dL^{(1)}}{dt} + a * L^{(1)} = b \tag{6}$$

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

$$\hat{a} = \begin{bmatrix} a \\ b \end{bmatrix} = (B^T B)^{-1} B^T Y_N \tag{7}$$

Where the matrices B and Y_N are defined as follows:

$$B = \begin{bmatrix} -z^{(1)}(2) & 1 \\ -z^{(1)}(3) & 1 \\ \dots & \dots \\ -z^{(1)}(n) & 1 \end{bmatrix} \tag{8}$$

$$Y_N = \begin{bmatrix} l^{(0)}(2) \\ l^{(0)}(3) \\ \dots \\ l^{(0)}(n) \end{bmatrix} \tag{9}$$

And the member $z^{(1)}(k)$ in matrix B is defined as follows:

$$z^{(1)}(k) = \frac{l^{(1)}(k) + l^{(1)}(k-1)}{2} \quad k = 2, 3, \dots, n \quad (10)$$

Then the response sequence of GM (1, 1), that is predicted sequence, can be computed with Equation (11):

$$\hat{l}^{(1)}(k + 1) = (l^{(0)}(1) - \frac{b}{a}) e^{-ak} + \frac{b}{a} \quad k = 1, 2, \dots, n \quad (11)$$

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

$$\hat{l}^{(0)}(k + 1) = \hat{l}^{(1)}(k + 1) - \hat{l}^{(1)}(k) \quad k = 1, 2, \dots, n \quad (12)$$

The Verhulst model is a type of single variable and second order differential equation, which is defined in Equation (13) as follows:

$$\frac{dL^{(1)}}{dt} + a * L^{(1)} = b * (L^{(1)})^2 \quad (13)$$

Where a and b can be acquired from the following Equation (14):

$$\hat{a} = \begin{bmatrix} a \\ b \end{bmatrix} = (B^T B)^{-1} B^T Y_N \quad (14)$$

Where the matrices B and Y_N are defined in Equation (15) and (16) separately:

$$B = \begin{bmatrix} -z^{(1)}(2) & (z^{(1)}(2))^2 \\ -z^{(1)}(3) & (z^{(1)}(3))^2 \\ \dots & \dots \\ -z^{(1)}(n) & (z^{(1)}(n))^2 \end{bmatrix} \quad (15)$$

$$Y_N = \begin{bmatrix} l^{(0)}(2) \\ l^{(0)}(3) \\ \dots \\ l^{(0)}(n) \end{bmatrix} \quad (16)$$

And the member $z^{(1)}(k)$ in matrix B is same to Equation (12), it is presented in Equation (17) again for convenience:

$$z^{(1)}(k) = \frac{l^{(1)}(k) + l^{(1)}(k-1)}{2} \quad k = 2, 3, \dots, n \quad (17)$$

For Verhulst model, its response sequence can be presented as follows:

$$\hat{l}^{(1)}(k + 1) = \frac{a * l^{(1)}(0)}{b * l^{(1)}(0) + (a - b * l^{(1)}(0)) e^{ak}} \quad k = 1, 2, \dots, n \quad (18)$$

As a matter of fact, one saturated sequence already has some degree of exponential form, so the original sequence can be regarded as $L^{(1)}$ directly. $L^{(0)}$ in Equation (16) should be the IAGO sequence of the original sequence. Therefore, for Verhulst model, the predicted values of the original sequence can be acquired with Equation (18).

3.4 Prediction procedure

Grey model GM (1, 1) and Verhulst are used to predict the error rate based on most recent three historic data. An example of practice opportunities is given for how to get subsequence from a raw sequence as follows:

[4, 3, 1, 0, 0, 4, 3, 1, 1, 2]

Subsequence [4, 3, 1] from the raw sequence is extracted to build the grey model for initial prediction of error rate. The subsequence window is shifted to new data in next step, that is, the next subsequence is [3, 1, 0]. With the new subsequence, another new grey model is constructed for the prediction in next step. Repeat the building grey model process until the last subsequence. The two learning curve modes in Figure 3 are one of possible forms about real learning curves. In fact, the variants of two learning curve modes are also practical in real learning. They are shown in Figure 5.

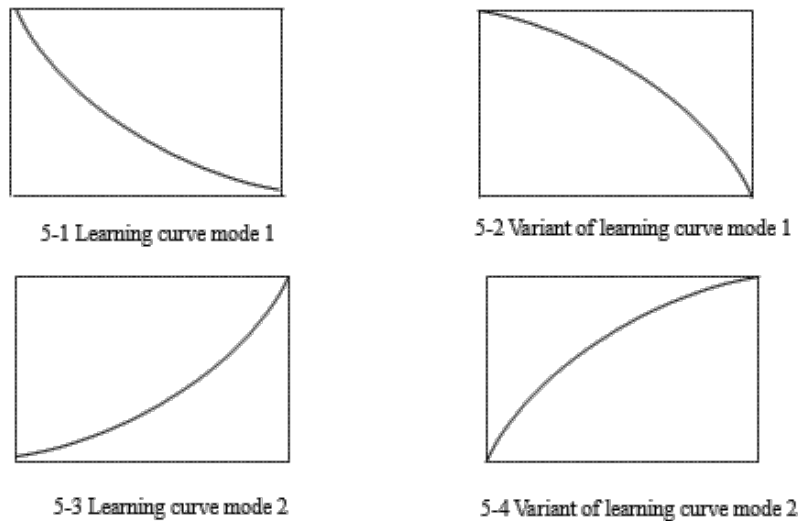


Fig. 5. Two learning modes and their variants

With the subsequence and determining its corresponding learning curve mode, the adequate grey model is chosen. The prediction procedure of grey model for the error rate of next practice opportunity is as follows:

Step 1: If a raw sequence length is equal or greater to 4, it is prepared to build the subsequence.

Step 2: Much unwanted fluctuation is contained in raw data, which shall be smoothed before prediction. Smooth technique reduces noise-based fluctuations in the sequences. The modified moving average of order 3 in the experiment is used to reduce fluctuation from the subsequence. A common smoothing method is the weighted moving average of order k as stated in [22]. A subsequence (l_1, l_2, l_3) is smoothly processed with the weighted moving average of order 3 as follow:

$$\frac{3 * l_1 + l_2}{4} \frac{l_1 + 2 * l_2 + l_3}{4} \frac{l_2 + 3 * l_3}{4} \tag{19}$$

Step 3: What learning curve mode about a subsequence can be determined by two grey modes in subsection 3.3. To predict the error rate, first thing, what learning curve mode of subsequence should be known. If the curve type of subsequence belongs to Figure 5-3 or 5-4, the corresponding grey model is GM (1, 1) or Verhulst separately. If the curve type of subsequence belongs to Figure 5-1 or 5-2, first do up flipping for them, the curve type will be converted to Figure 5-4 or 5-3. Then, use this up flipping subsequence to build grey model for the prediction of error rate. For those subsequences belonging to learning curve mode Figure 5-1 and 5-2, the current result of prediction is not the real result of original subsequence. The expected prediction is acquired with doing down flipping again.

Step 4: In this step, build grey models based on the curve type of subsequence and predict the error rate in next practice opportunity until the sequence ends.

4 Experiment

Dataset “Geometry Area (1996-97)” in DataShop is used in the experiment. The number of student is 59. Total number of transactions is 6,778. Total number of KC models is 50, in which 10 KC models are randomly chosen in the experiment.

For each prediction, the predicted error rate of practice opportunity based on two grey models is compared with the predicted error rate based on LFA model. For analysis of two prediction methods’ pros and cons, root mean square error (RMSE) is used as evaluation measure. Better the degree of fitting real value is, lower RMSE is. RMSE is computed as follows:

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (p_i - \hat{p}_i)^2}{n}} \tag{20}$$

Where \hat{p}_i is the predicted value of error rate of practice opportunity on original value of error rate p_i , which is error rate for all students’ first attempt (i.e., students’ incorrect answer or hint), and n is number of practice opportunity for target KC. In Table 1, it is RMSE summary for randomly chosen 10 KC models. Each KC model consists of some KCs. RMSE of grey models is much better than LFA. It shows that grey models have good capacity to predict the error rate of student performance. LFA is successfully used to redesign KCs for all students practicing reasonable opportunities by identifying hidden KCs or poorly defined KCs, on the whole, students use less time to finish learning KC, additionally, predicted learning curve based on LFA also indicates if students have effective learning as stated in [23].

However, in Figure 2-3 and 2-6, two real learning curves of individual student compared with predicted learning curves based on LFA presents to largely fluctuate. Individual student apparently cannot get the better recommendation on individual practicing opportunities for the mastery of target KC with LFA prediction.

Table 1. RMSE values comparison between LFA and grey model for KC models

KC Models	Number of KC	RMSE of LFA	RMSE of Grey Models
Area	2	0.4274	0.1597
Circle-Collapse	8	0.4058	0.3091
Concept	9	0.4089	0.3091
Decompose	13	0.4048	0.1433
Geometry	1	0.4274	0.1523
Original	15	0.4072	0.0648
Single-KC	1	0.4285	0.1517
Single Plus Unique	133	0.4534	0.1519
Textbook	12	0.4048	0.1560
Unique-step	132	0.4008	0.0732

In Table 1, RMSE of grey models is much better than RMSE of LFA, it seems to provide better predicting of practice opportunities for KC mastery. In Figure 6, x axis is number of practice opportunity, y axis is student performance. In Figure 6, it shows that predicted learning curves based two prediction methods and real learning curves about 10 KC models in Table 1.

It seems to totally overlay between real learning curves in blue and predicted learning curves in green based on the grey models. Once such charts in Figure 6 expand, it shows the predicted learning curves based on the grey models are often lagged in some practice opportunities compared with real learning curves. This kind of lagging occurs in inflection point of real learning curve.

The grey models in subsection 3.3 work well if learning curves are only down or up tendency. In subsection 3.4, grey models' computing based on moving windows technique only adapts to the same tendency change in real learning curve for predicting error rate in next practice opportunity. However, once inflection point occurs in real learning curve, grey models are defeated as they do not have any information about practices of inflection points, such as a practice crossing how many KCs, representation method of practice et al.

If such information is fed into grey models in advance, grey models' computing based on moving windows technique shall adjust its direction in some degree, and it may improve the prediction for the error rate of next practice opportunity in real learning curve.

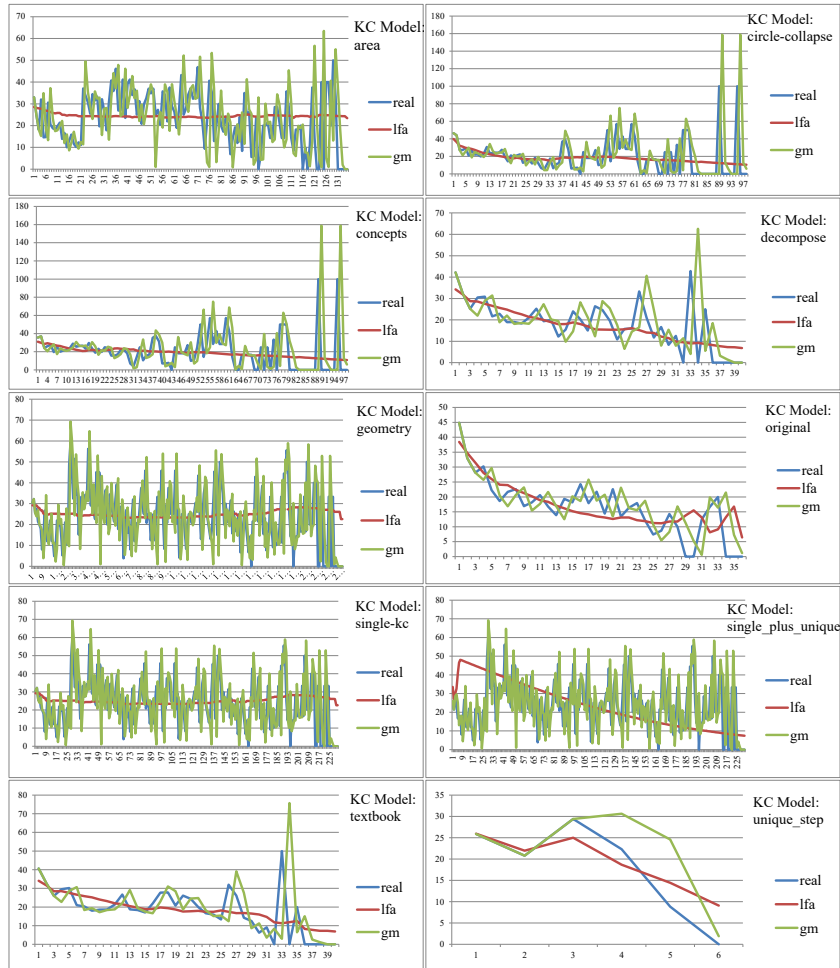


Fig. 6. Predicted learning curves in green based on grey models and in red based on LFA

5 Conclusion

In this paper, the two grey models are used to predict the error rate of student's each practice opportunity for target KC. The experimental result shows that the prediction of proposed method based on the grey models is much better than LFA. It reveals the proposed method has capacity to recommend the number of practice opportunity for individual student. However, the defect of one inflection point in real learning curve exists in the current version of proposed method, and its initial possible solution is discussed. In future, how to acquire and evaluate information about KC practices of inflection points in real learning curves shall be studied for individual student personalized tutoring in ITS.

6 Acknowledgement

This work is supported by Xi'an University of Posts and Telecommunications project (103-204020217) and Association of Fundamental Computing Education in Chinese Universities project (2018-AFCEC-360). We used the 'Geometry Area (1996-97)' dataset accessed via DataShop (pslcdatashop.org).

7 References

- [1] Vinicius Gottin, Haydée Jiménez and Anna Carolina Finamore, An analysis of Degree Curricula through Mining Student Records. IEEE 17th International Conference on Advanced Learning Technologies (ICALT 2017), Politehnica University of Timisoara, July 3-7 2017, Timisoara, Romania. pp. 276-280. <https://doi.org/10.1109/ICALT.2017.54>
- [2] Jianhui Chen and Jing Zhao (2018). An Educational Data Mining Model for Supervision of Network Learning Process. International journal of Emerging Technologies in Learning, 13(11): pp. 67-76. <https://doi.org/10.3991/ijet.v13i11.9599>
- [3] Xiaowei Hao and Shanshan Han (2018). An Algorithm for Generating a Recommended Rule Set Based on Learner's Browse Interest. International journal of Emerging Technologies in Learning, 13(11): pp. 102-115. <https://doi.org/10.3991/ijet.v13i11.9604>
- [4] Tung-Cheng Hsieh and Tzone-I Wang (2010). A Mining-based Approach on Discovering Course Pattern for Constructing Suitable Path. Expert Systems with Applications, 37: pp. 4156-4167. <https://doi.org/10.1016/j.eswa.2009.11.007>
- [5] Gurhan Durak and Serkan Cankaya (2018). The Current State of The Art in Learning Space: A Systematic Review Study. International journal of Emerging Technologies in Learning, 13(11): pp. 208-223. <https://doi.org/10.3991/ijet.v13i11.9247>
- [6] Michel C. Desmarais and Ryan S. J. d. Baker (2012). A Review of Recent Advances in Learner and Skill Modeling in Intelligent Learning Environments. User Model User-Adaptation Interaction, 22: pp. 9–38. <https://doi.org/10.1007/s11257-011-9106-8>
- [7] Henrique Seffrin, Ig I. Bittencourt, Seiji Isotani and Patricia A. Jaques, Modelling Students' Algebraic Knowledge with Dynamic Bayesian Networks. IEEE 16th International Conferences on Advanced Learning Technologies (ICALT 2016), July 25-28 2016, Austin, TX, USA. pp. 44-48. <https://doi.org/10.1109/ICALT.2016.96>
- [8] Xingliang Chen, Antonjia Mitrovic and Moffat Mathews, How much Learning Support Should be Provided to Novice and Advanced Students. IEEE 17th International Conferences on Advanced Learning Technologies (ICALT 2017), July 3-7 2017, Timisoara, Romania. pp. 39-43. <https://doi.org/10.1109/ICALT.2017.43>
- [9] Albert T. Corbett and John R. Anderson (1995). Knowledge Tracing: Modeling the Acquisition of Procedural Knowledge. User Model User-Adaptation Interaction, 4: 253-278. <https://doi.org/10.1007/BF01099821>
- [10] Olle Bälter, Dawn Zimmaro and Candace Thille (2018). Estimating the minimum number of opportunities needed for all students to achieve predicted mastery. Smart Learning Environments, 5(1): 15. <https://doi.org/10.1186/s40561-018-0064-z>
- [11] Benjamin S. Bloom (1968). Learning for Mastery. Instruction and Curriculum. Regional Education Laboratory for the Carolinas and Virginia, Topical Papers and Reprints, Number 1. Evaluation Comment, 1(2): pp. 1-12.
- [12] Newell A. and Rosenbloom P. (1981). Mechanisms of Skill Acquisition and the Law of Practice. In Anderson J. (ed.): Cognitive Skills and their Acquisition, Hillsdale, NJ, Lawrence Erlbaum Associates.

- [13] Panagiotis L., Diomidis S. and Vasileios V. (2008). Power Laws in Software. *ACM Transactions on Software Engineering and Methodology*, 18(1): pp. 1-26. <https://doi.org/10.1145/1391984.1391986>
- [14] Seibel R. (1963). Discrimination reaction time for a 1,023 alternative task. *Journal of Experimental Psychology*, 66(3): pp. 215-226.
- [15] Ritter, F.E. and Schooler, L. J. (2001). 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. <https://doi.org/10.1016/B0-08-043076-7/01480-7>
- [16] Kenneth R. K. and Santosh M., Distinguishing Qualitatively Different Kinds of Learning Using Log Files and Learning Curves. *Log Analysis Workshop in 7th International Conference of Intelligent Tutoring Systems (ITS 2004)*, August 30- September 3, Maccio, Alagoas, Brazil.
- [17] Cen H., Koedinger K., and Junker B., Comparing Two IRT Models for Conjunctive Skills. *9th international Conference on intelligent Tutoring Systems (ITS 2008)*, June 23-27, Montreal, Canada. pp. 796-798. https://doi.org/10.1007/978-3-540-69132-7_111
- [18] Cen H., Koedinger K. and Junker B., Learning Factors Analysis – A General Method for Cognitive Model Evaluation and Improvement. *8th International Conference on Intelligent Tutoring Systems (ITS 2006)*, Jhongli, Taiwan, China. pp. 164-175. https://doi.org/10.1007/11774303_17
- [19] DataShop, <https://pslclatashop.web.cmu.edu/index.jsp>, 2018.9.
- [20] Tzu-Yi P., Keisuke H., Hsin-Hsien H. and Chun-Ming Hsieh (2007). Using grey system theory to evaluate transportation effects on air quality trends in Japan. *Transportation Research Part D: Transport and Environment*, 12 (3): pp. 158-166. <https://doi.org/10.1016/j.trd.2007.01.007>
- [21] J. Deng (2008). Control Problems of grey systems. *System Control Letter*, 1(5): pp. 288-294. [https://doi.org/10.1016/S0167-6911\(82\)80025-X](https://doi.org/10.1016/S0167-6911(82)80025-X)
- [22] J. W. Han and M. Kamber (2001). *Data Mining: Concepts and Techniques*. San Fransisco, CA: Morgan Kaufmann.
- [23] Kenneth R. Koedinger, John C. Stamper, Elizabeth A. McLaughlin and Tristan Nixon, Using Data-Driven Discovery of Better Student Models to Improve Student Learning. *International Conference on Artificial Intelligence in Education (AIED 2013)*, July 9-13, Memphis, TN, USA. pp. 421-430. https://doi.org/10.1007/978-3-642-39112-5_43

8 Authors

Manqiang Liu is with the College of Electrical and Information Engineering, Lanzhou University of Technology, Lanzhou, China. His research interests include industrial equipment diagnosis, data mining, and learning analytics. He has rich engineering experiences in the designing, development and onsite debugging and diagnosis of non-ferrous metal equipment control (e-mail: liumq2004@163.com).

Qingsheng Zhang is with Xi'an University of Posts and Telecommunications, Xi'an, China. His research interests include ubiquitous computing, context-aware computing applications, distributed computing and applications, data mining, and adaptive eLearning. He has published several papers in International conferences and journals (email: zhangqs@xupt.edu.cn).

Article submitted 2018-11-19. Resubmitted 2019-03-28. Final acceptance 2019-03-29. Final version published as submitted by the authors.