Traditional Knowledge Tracing Models for Clustered Students

Deliang Wang[1], Zhi Zhang[1], Jiachen Song[2], Yu Lu[1,2,*]

[1]School of Educational Technology, Faculty of Education, Beijing Normal University, Beijing, China.
[2]Advanced Innovation Center for Future Education, Beijing Normal University, Beijing, China.

Abstract

Against the background of the worldwide COVID-19 pandemic, online learning has currently become one of the dominant educational forms. For more effective online learning, knowledge tracing that can dynamically estimate the knowledge state of learners should be paid more attention. This study gives an overall introduction of knowledge tracing models. We illustrate the development of the Bayesian Knowledge Tracing model and clarify its concrete mathematic principle. In addition, an individualized method based on clustered students for the Bayesian Knowledge Tracing model is initially proposed, which changes the individualization level from a group of all students to subgroups. To confirm whether this individualized method can be generalized to other knowledge tracing models, we also test it on logistic knowledge tracing models. Therefore, we provide an introduction about the principles of three logistic knowledge tracing models. We evaluate our method on the four models with two international public educational datasets. The results show that all knowledge tracing models for clustered students outperform the original models without clustering. The Bayesian Knowledge Tracing model and item response theory for clustered students both gain good improvement in their performance, while the Performance Factors Analysis for clustered students has only few improvements. The reason for the experimental results is discussed at the end. Overall, this paper provides a general method for further promoting the performance of traditional knowledge tracing models that requires more studies.

Keywords

Bayesian Knowledge Tracing Model, Learning Factors Analysis, Performance Factors Analysis, Clustered Students

1. Introduction

The COVID-19 pandemic began at the end of 2019 and soon swept across the world. To prevent it spreading to campuses, most countries cancelled or delayed the new semester in 2020. UNESCO estimates that approximately 1.3 billion students are currently in the state of suspended classes. To continue learning activities, many classes have been moved to the Internet environment by leveraging different online learning platforms. Therefore, it is most likely that online learning will become one of the major ways for students to learn even after COVID-19 ends. However, some issues have been found that prevent effective online learning. First, online learning activities are easily separated by time and space; thus, teachers find it difficult to timely receive feedback from their students behind the screens. This causes problems for learning and teaching to promote one another (Yu &Wang, 2020). Second, many teachers who lack

information and communications technology (ICT) skills might encounter difficulties in choosing proper internet resources and platforms, which results in their students being unable to get personalized learning support (Wang, Wei, & Zong, 2020). One possible solution to strengthen online education is to automatically obtain an accurate understanding of students driven by their learning data (Song, Xu, & Li, 2020; Zhong, 2020).

Among the tasks of understanding students, the most basic one is to estimate their mastery levels of knowledge. If such estimation can be conducted, then online learning platforms can be designed to recommend appropriate learning resources and help students achieve learning goals much more efficiently. Based on the estimation results, the online platforms could also conduct predictions on their future performance and accordingly, sound early alarms, if necessary. In the context of student modelling, the process of estimating a student's mastery level of knowledge is called knowledge tracing (KT). Against the background of COVID-19 that has caused online education to become the dominant learning and teaching form, KT models should be paid more attention to, and further studies on them should be conducted. This paper gives an overall introduction of KT models and describes the principles of some traditional KT models, such as the Bayesian Knowledge Tracing model, Learning Factors Analysis and Performance Factors Analysis. We propose a generalized method that can promote the performance of most traditional KT models, which can serve as a basis for the further promotion of traditional KT models' performance.

2. Related Work

Knowledge tracing models can be classified into two types. The first type is the traditional KT model, including the Bayesian Knowledge Tracing model (Corbett & Anderson, 1994), Learning Factors Analysis (Cen, Koedinger, & Junker, 2006), Performance Factors Analysis (Pavlik, Cen, & Koedinger, 2009), etc. The second type is Deep Learning-based Knowledge Tracing (DLKT) models, such as Deep Knowledge Tracing (Piech et al., 2015) and the Dynamic Key Value Memory Network (Zhang, King, & Yeung, 2017). The two types both have strengths and weaknesses. Traditional KT models have an average performance level, but they are highly interpretable. Based on the fact that they can satisfy the basic needs of students and teachers, such KT models can be applied to various types of learning platforms. DLKT models have a superior performance most of the time and sometimes, their performance is inferior or similar to traditional KT models, such as the modified BKT in Khaja's research (Khajah, Lindsey, & Mozer, 2016). However, because of the mathematic models that they are based on, they are highly uninterpretable. DLKT models output a probability, which informs students of a probability that they can answer a question correctly. However, these models do not give the reason for what occurs inside the model and why these models make such predictions. If they make an incorrect prediction and accordingly recommend unsuitable learning resources, then they will not promote learning efficiency and will instead waste students' time, contrary to the initial expectations. The uninterpretable feature inhibits their application in education. Therefore, there are two trends in the KT domain, including promoting the performance of KT models and explaining DLKT models. In this paper, we focus on the promotion of KT models' performance and take the Bayesian Knowledge Tracing model and logistic knowledge tracing models as examples.

The standard Bayesian Knowledge Tracing (BKT) model was proposed by Corbett and Anderson to estimate the mastery level of students and predict their future performance by allocating four parameters for each skill, including the probability of initial mastery, the ability of learning, the probability of guessing and slipping. Based on the estimation and prediction, the learning platform can be designed to recommend corresponding guidance for learners until they can reach a master level. However, BKT is skill-oriented, which means that all learners share a set of parameters for one skill. The weakness is that learners who fall above average receive a large number of exercises after they have mastered the skill, while learners below the average level cannot receive adequate resources although they have not reached the master level. The ideal model is that every learner has one set of parameters for each skill, but the data cannot satisfy this need even in the era of big data. Subsequently, many scholars extend the standard BKT model to make it more individualized. Lee and Brunskill modified BKT into a learner-oriented model (Lee & Brunskill, 2012). Learner-oriented means that each learner has one set of parameters for all skills, which is different from standard BKT. The results show that the number of practice opportunities arranged for each learner changes and becomes much more suitable for their knowledge state. Yudelson, Koedinger, and Gordon individualized the parameter of initial mastery and learning ability by dividing them into a skill part and a learner part (Yudelson, Koedinger, & Gordon, 2013). Additionally, the modified BKT model outperforms the standard BKT model, with a higher accuracy.

Except for modifications in individualization, scholars also extend BKT from other angles such as parameter issues and reality issues. In terms of parameter issues, Nelimarkka found that setting the parameter of initial mastery too high causes more errors in other parameters (Nelimarkka & Ghori, 2014). For example, if we set the initial mastery to 0.9, then the limited data cannot support us and will generate reliable estimates for the parameters of guessing and slipping. Therefore, BKT has to make some changes when facing learners with a high knowledge base. Meanwhile, the parame-

ters of guessing and slipping cannot be set high because such settings will cause BKT to degrade (Baker, Corbett, & Aleven, 2008). Specifically, when a learner answers a question correctly (or incorrectly), we cannot tell whether it is because the learner has mastered (or not mastered) this skill or whether the learner guesses correctly (or slips). In terms of model reality issues, which means making BKT closer to the complicated fact, scholars incorporate other information in BKT. Klingler pointed out that BKT should consider the relationship between skills, and the accuracy becomes higher because it is likely that one question contains multiple skills and because the mastery of one skill affects the mastery of its related skills (Käser, Klingler, Schwing, & Gorss, 2014). Pardos and Heffernan introduced the difficulty of the questions into BKT (Pardos & Heffernan, 2011). The increase in this difficulty causes learners to give incorrect answers even if they have mastered the corresponding skill. In addition, BKT performs better than the standard BKT after incorporating this difficulty. Wang and Hefferman suggested changing the binary representation of the correctness into a consecutive representation between 0 and 1, and their experimental results also show better performance (Wang & Heffernan, 2013).

3. Traditional Knowledge Tracing Models

Traditional knowledge tracing models can be divided into two types. The first type is the Bayesian Knowledge Tracing model, and the second type contains logistic knowledge tracing models, mainly including Learning Factors Analysis and Performance Factors Analysis. In this section, we explain the mathematic principles of these KT models.

3.1. Bayesian Knowledge Tracing Model

BKT belongs to a Hidden Markov Model (HMM). It divides learners' knowledge state into observable responses and an unobservable actual mastery level. Based on the item response on a skill, BKT can estimate the learner's mastery level using the Bayesian formula and then make a prediction on the learner's performance of the same skill. Specifically, BKT models the knowledge state of students on each skill in a binary form, namely, "mastered" or "unmastered". The unmastered state can be transformed to the mastered state through learning or practicing, and vice versa, the mastered state can be transformed to the unmastered state through forgetting, while standard BKT assumes that there is no forgetting process during learning, which is not practical (Qiu, Lu, Pardos, & Heffernan, 2011). Here, we add the forgetting process to BKT. The correctness of an answer is also binary, specifically, correct or incorrect. When in the mastered state, the student may give incorrect answers by slipping, and when in the unmastered state, it is possible to answer the question correctly by guessing. Apart from learning ability P(T), forgetting P(F), slipping P(S) and guessing P(G), BKT also has one more parameter: the probability of mastering a skill initially $P(L_0)$. The interpretation of the parameters can be seen in Table 1. They are trained through the expectation–maximization (EM) algorithm. We also show the state transition matrix and emission matrix of BKT in Tables 2 and 3, respectively.

Table 1. Notations and explanations of parameters in standard BKT

Parameter	Explanation
P(L₀)	Before learning or practicing at time t, the probability of mastering Skill K for Learner i: $P(K_i^0=1)$
P(T)	Learner i transforms the unmastered state on Skill K at time n into the mastered state at time n+1 through learning or practicing: $P(K_i^{n+1}=1\mid K_i^n = 0)$
P(S)	Learner i gives an incorrect answer by slipping when in the mastered state on Skill K at time n: $P(O_n=0\mid K_i^n = 1)$
P(G)	Learner i gives a correct answer by guessing when in the unmastered state on Skill K at time n: $P(O_n=1\mid K_i^n = 0)$
P(F)	Learner i transforms the mastered state on Skill K at time n into the unmastered state at time n+1 through forgetting: $P(K_i^{n+1}=0\mid K_i^n = 1)$

Table 2. The state transition matrix in standard BKT

	Unmastered	Mastered
Unmastered	1-P(T)	P(T)
Mastered	P(F)	1-P(F)

Table 3. The emission matrix in standard BKT

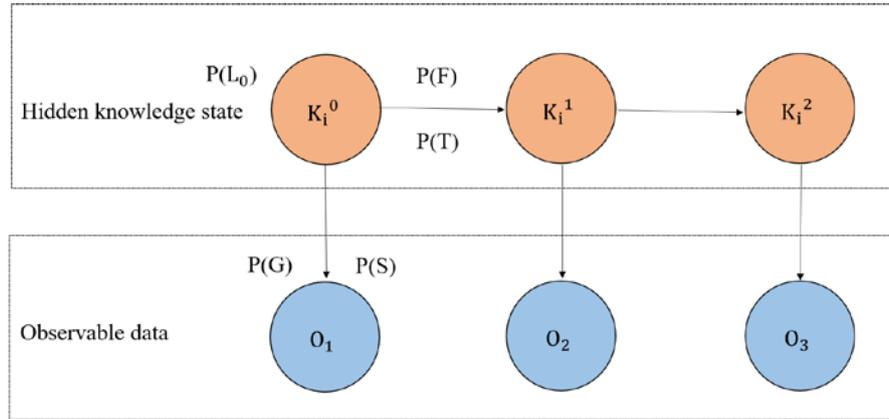|	Incorrect answer	Correct answer
Unmastered	1-P(G)	P(G)
Mastered	P(S)	1-P(S)

Figure 1. The working mechanism of standard BKT.

The working mechanism of BKT is shown in Figure 1 and can be represented by the two phases of estimating the knowledge state and predicting future performance. Based on learners' item response on a skill, the Bayesian formula is used to update the mastery level for this skill. For a correct answer, the mastery level is calculated through equation (1). For an incorrect answer, the mastery level is computed through equation (2).

$$P(K_i{}^n = 1|O_n = 1) = \frac{P(K_i{}^n=1)*(1-P(S))}{P(K_i{}^n=1)*(1-P(S))+\left(1-P(K_i{}^n=1)\right)*P(G)} \tag{1}$$

$$P(K_i{}^n = 1|O_n = 0) = \frac{P(K_i{}^n=1)*P(S)}{P(K_i{}^n=1)*(P(S))+\left(1-P(K_i{}^n=1)\right)*(1-P(G))} \tag{2}$$

After estimation, we use the total probability rule to update the corresponding initial knowledge state at the next time step by combing learning and not forgetting.

$$P\left(K_i{}^{n+1} = 1\right) = P(K_i{}^n = 1|O_n) * (1 - P(F)) + (1 - P(K_i{}^n = 1|O_n)) * P(T)) \tag{3}$$

Then, we make predictions on students' performance. When confronted with an item on the same skill, the learner may give a correct answer by guessing correctly or not slipping. Meanwhile, he or she may also give an incorrect answer by guessing incorrectly or slipping. The two scenarios are as follows.

$$P(O_{n+1} = 1) = P(K_i{}^{n+1} = 1) * (1 - P(S)) + (1 - P(K_i{}^{n+1} = 1)) * P(G) \tag{4}$$

$$P(O_{n+1} = 0) = P(K_i{}^{n+1} = 1) * P(S) + (1 - P(K_i{}^{n+1} = 1)) * (1 - P(G)) \tag{5}$$

3.2. Logistic Knowledge Tracing Models

3.2.1. Item Response Theory

Since the 1950s, item response theory (IRT) has been widely applied in educational tests. It is used to illustrate the relationship between learners' responses on items and their abilities (Rasch, 1993). Specifically, based on the differentiating level of an item, the difficulty of an item and the ability of a learner, the two-parameter IRT model outputs the probability of answering this item properly, whose mathematical expression follows. Except for predicting students' response, IRT can also be used to evaluate the item's difficulty level and each student's ability. Strictly, IRT does not belong to KT domain because it assumes that the knowledge state of learners does not change. As the basis of logistic models for knowledge tracing, we provide a brief introduction of IRT here and use it as our baseline for the experiment.

$$P\left(\alpha_j, \theta_i, \beta_j\right) = \frac{1}{1+e^{-\alpha_j(\theta_i-\beta_j)}} \tag{6}$$

3.2.2. Learning Factors Analysis

Learning Factors Analysis (LFA) is a semi-automated method for improving a cognitive model that consists of a statistical model, human expertise and a combinatorial search. Its statistical model is essentially a KT model. LFA originates from the learning curve, which found a power relationship between the error rate of performance and the amount of practice. However, the learning curve does not take into account the facts that there are many learners and that a question may contain various skills. Therefore, LFA adapts and extends the learning curve.

$$\ln\left(\frac{p}{1-p}\right) = \alpha_i X_i + \sum_j \beta_j Y_j + \sum_j \gamma_j Y_j T_j \tag{7}$$

X_i and Y_j stand for the learner i and skill j, respectively. α_i means the initial knowledge that learner i has, which indicates that different students may have prior knowledge. β_j is the level of skill j known by students before learning, assuming that some skills have already been more or less known. γ_j represents the learning rate of skill j, which reflects the difficulty level. The summation sign signifies the number of skills involved in the question. With the accumulation of learning on all skills, we can obtain the probability p that a learner will answer an item correctly. LFA thus inherits student ability and skill difficulty in IRT, incorporates the learning curve and also extends to the scenario of multiple skills in one item.

3.2.3. Performance Factors Analysis

Performance Factors Analysis (PFA) is an adaptation of LFA. LFA is sensitive to the practice times but ignores the correct and incorrect responses produced by learners. PFA assumes that each learner is unique; therefore, his or her learning is not accumulated only by practice frequency in an identical fashion. Each learner also learns through his or her own correct and incorrect trials. Therefore, PFA makes the following adaptation based on LFA.

$$\ln\left(\frac{p}{1-p}\right) = \sum_{j \in KCs}(\beta_j Y_j + \gamma_j Y_j s_{ij} + \rho_j Y_j f_{ij}) \tag{8}$$

Y_j stands for skill j included in one item. β_j has the same meaning as it has in LFA, which is the level of skill j known by students before learning. s_{ij} and f_{ij} represent the number of prior successes and failures, respectively, of learner i for skill j. γ_j and ρ_j are the corresponding weights of s_{ij} and f_{ij}, which indicate the importance of successful and unsuccessful trials for learning. By summing all accumulated learning of skills in an item, PFA can calculate the probability p of solving the item correctly.

4. Method

As mentioned earlier, BKT is skill-oriented, and the set of parameters for each skill is average, which cannot differentiate learners with different mastery levels, especially learners above or below the average. Here, we propose a method that can distinguish students of different levels and promote the performance of the model. Meanwhile, this individualized method can be generalized to other KT models.

For BKT, it is not possible to allocate with a set of parameters each student to each skill because the number of items that each learner interacts with in the system is limited. Therefore, we change the level of individualization from each person to a group. We propose to divide all learners into different groups based on their ability at something, such as their learning style or learning ability. Knowledge tracing essentially models students' cognition; thus, it is worth attempting to determine whether modifying BKT for different cognitive levels can make it outperform standard BKT or not. However, it is difficult to collect massive data about students' cognitive information. Given the data that we can access from international public datasets, we propose to divide students into three clusters based on their correct rate on all items and regard it as a symbol of their learning ability. The threshold of the correct rate is set to 0.8 and 0.6. A learner with a correct rate greater than 0.8 is considered above the average level and a learner with a correct rate less than 0.6 is regarded as below the average level. Learners whose correct rate is between 0.6 and 0.8 belong to the average group. We also use this individualization method on the logistic KT models.

5. Experiment

5.1. Database

We choose two international public educational datasets to evaluate the KT models for clustered students, including ASSISTment 2009 and ASSISTment 2015. They are provided by the ASSISTment online tutoring system and are often adopted for KT tasks. Both of them contain the interactions between learners and items for different skills, including the learner id, skill id, skill name, and the correctness of each interaction. Detailed information about the two datasets can be seen in Table 4.

- ASSISTment 2009: We take its updated skill builder dataset for math. All the interactions without labeling skills or skill names are filtered out. The dataset contains 325,637 interactions with 4,151 learners for 110 skills.
- ASSISTment 2015: The correctness of each interaction is not always binary, which needs to be modified for the KT models. We set the correctness greater than 0.5 to 1 and the correctness less than 0.5 to 0. The dataset contains 708,631 interactions with 19,917 learners for 100 skills.

5.2. Model Training and Baselines

We divide each dataset into three clusters based on the correct rate of learners. The cluster information can be seen in Tables 5 and 6. We select 80% of the data randomly as the training data and the remaining 20% as the test data. When training BKT for student clusters, we repeat the experiment 20 times, considering that the random choice of the initial parameters may cause the expectation–maximization algorithm to be stuck on the local optimal result.

Table 4. Basic information of the two databases

Dataset	#Learners	#Skills	#Interactions
ASSISTment 2009	4,151	110	325,637
ASSISTment 2015	19,917	100	708,631

Table 5. Cluster information of ASSISTment 2009

ASSISTment 2009	#Learners	#Skills	#Interactions
Cluster 1	1,149	104	62,070
Cluster 2	1,394	110	168,296
Cluster 3	1,608	110	95,271

Table 6. Cluster information of ASSISTment 2015

ASSISTment 2009	#Learners	#Skills	#Interactions
Cluster 1	8,404	100	224,413
Cluster 2	7,824	100	376,281
Cluster 3	3,689	100	107,937

Since the KT task here is a binary classification problem (a correct answer or an incorrect answer), we use the area under a curve (AUC) as our metric. Meanwhile, the accuracy metric is also adopted. To confirm whether the cluster method can be generalized to other traditional KT models, we also conduct experiments on the logistic KT models, including the three models that we introduced in the previous section.

6. Results and Discussion

Table 7 shows the experimental results between the original KT models and the KT models for clustered students (symbolized as c-KT). We can see c-BKT gain a performance improvement of 2.9% in accuracy and 5.8% in the AUC in ASSISTment 2009 and an improvement of 2.04% in accuracy and 1.8% in the AUC in ASSISTment 2015. For the logistic KT models, c-IRT, c-LFA, and c-PFA all outperform the original IRT, LFA and PFA. Surprisingly, the c-IRT model receives the most improvement in both datasets, which is 6.13% in accuracy and 12.3% in the AUC in ASSISTment 2009 and 1.1% in accuracy and 7.1% in the AUC in ASSISTment 2015. There are only a few improvements in the performance of c-PFA. In ASSISTment 2015, c-PFA improves only in the AUC, and the accuracy even decreases.

Table 7. The performance of the KT models in two datasets

Model	ASSISTment2009		ASSISTment2015	
	ACC	AUC	ACC	AUC
BKT	67.92%	0.660	73.62%	0.652
c-BKT	70.82%	0.718	75.66%	0.670
IRT	65.68%	0.611	73.00%	0.632
c-IRT	71.81%	0.734	74.11%	0.703
LFA	65.32%	0.560	72.89%	0.583
c-LFA	66.81%	0.629	73.00%	0.632
PFA	66.44%	0.624	73.31%	0.643
c-PFA	68.77%	0.640	73.25%	0.676

Notes: c-KT means KT models for clustered students.

As mentioned previously, it is highly possible that BKT will gain an improvement in its performance because learners can receive a set of parameters that better fits their mastery level after clustering.

For the three logistic models, the original IRT and LFA are more individualized than PFA because IRT and PFA both have a parameter that indicates learner ability or the initial knowledge. However, the skill parameters for IRT are obtained from all the students, which suggests that the average level cannot fit learners above or below the average level well. This is why c-IRT considerably outperforms IRT. Although PFA does not contain a parameter for learners, it takes their individualized learning trajectories into account and treats correct trials and incorrect trials differently. Moreover, the experimental results validate the individualized assumption of PFA.

7. Conclusion

This study gives an overall introduction of knowledge tracing models against the background of the COVID-19 pandemic, which has spread all over the world and caused online learning to become the dominant educational form. We explain the development of the Bayesian Knowledge Tracing model and clarify its concrete mathematic principle. In this paper, we propose an individualized method based on clustered students for BKT initially and change the individualization level from all persons to groups. To confirm whether this individualized method can be generalized to other KT models, we also test it on logistic KT models. Therefore, we also provide an introduction about the principles of three logistic KT models. We evaluate our method on the four KT models with two international public educational datasets. All KT models for clustered students outperform the original KT models, among which BKT and IRT gain the most improvement in performance, while PFA has few improvements. We infer that these results are related to the educational assumptions within each model. Overall, this paper provides a basis for the further promotion of the performance of traditional KT models by changing the individualization level.

Acknowledgements

References

Baker, R. S., Corbett, A. T., & Aleven, V. (2008). More Accurate Student Modeling through Contextual Estimation of Slip and Guess Probabilities in Bayesian Knowledge Tracing. In *Proceeding of the 9th International Conference on Intelligent Tutoring Systems* (pp. 406-415). Berlin: Springer.

Cen, H., Koedinger, K., & Junker, B. (2006). Learning Factors Analysis – a General Method for Cognitive Model Evaluation and Improvement. In *Proceedings of the 9th International Conference on Intelligent Tutoring Systems* (pp. 64-175). Berlin: Springer.

Corbett, A. T., Anderson, J. R. (1994). Knowledge Tracing: Modeling the Acquisition of Procedural Knowledge. *User Modeling & User Adapted Interaction, 4*(4), 253-278. doi.org/10.1007/BF01099821.

Käser, T., Klingler, S., Schwing, A. G., & Gorss, M. (2014). Beyond Knowledge Tracing: Modeling Skill Topologies with Bayesian Networks. In *Proceeding of the 12th International Conference on Intelligent Tutoring Systems* (pp. 188-198). Honolulu: Springer.

Khajah, M., Lindsey, R., & Mozer, M. (2016). How deep is knowledge tracing? In *Proceedings of the 9th International Conference on Educational Data Mining* (pp. 94-101). North Carolina: International Educational Data Mining Society.

Lee, J. I., Brunskill. E. (2012). The Impact on Individualizing Student Models on Necessary Practice Opportunities. In *Proceedings of the 5th International Conference on Educational Data Mining* (pp. 910-915). Chania: www.educationaldatamining.org.

Montavon, G., Samek, W., & Müller, K. (2017). Methods for interpreting and understanding deep neural networks. *Digital Signal Processing, 73*, 1-15. doi.org/10.1016/j.dsp.2017.10.011.

Nelimarkka, M., Ghori, M. (2014). The Effect of Variations of Prior on Knowledge Tracing. In *CEUR Workshop Proceeding on Educational Data Mining* (pp. 146-150). London: CEUR-WS.

Pardos, Z. A., & Heffernan, N. T. (2011). KT-IDEM: Introducing Item Difficulty to the Knowledge Tracing Model. In *Proceeding of 19th International Conference on User Modeling, Adaption and Personalization* (pp. 243-254). Berlin: Springer.

Pavlik, P. I., Cen, H., & Koedinger, K. R. (2009). Performance Factors Analysis—A New Alternative to Knowledge Tracing. *Proceedings of the 14th International Conference on Artificial Intelligence in Education* (pp. 531-538). Amsterdam, Netherlands: IOS Press.

Piech, C., Spencer, J., Huang, J., Ganguli, S., Sahami, M., Guibas, M., & Guibas, L. (2015). Deep Knowledge Tracing. In *Advances*

in Neural Information Processing Systems 28 (pp. 505-513). Montreal: Curran Associates, Inc.

Qiu, Y., Lu, H., Pardos, Z. A., & Heffernan, N. T. (2011). Does Time Matter? Modeling the Effect of Time with Bayesian Knowledge Tracing. In *Proceedings of the 4th International Conference on Educational Data Mining* (pp. 139-148). Eindhoven*:* www.educationaldatamining.org.

Rasch, G. (1993). *Probabilistic Models for some Intelligence and Attainment Tests*. Chicago, IL: MESA Press.

Song, L., Xu, L., & Li, Y. (2020). Precision 0nline Teaching + Home Study Model: A Feasible Way to Improve the Quality of Study for Students during Epidemic. *China Educational Technology, 3*, 114-122. doi:10.3969/j.issn.1006-9860.2020.03.016.

Wan, H., Wang, D. (2016). Investigation on Adaptive Learning Mechanism of Big Data based on Knewton Platform. *Modern Educational Technology, 26*(05), 5-11. doi:10.3969/j.issn.1009-8097.2016.05.001.

Wang, J., Wei, Y., & Zong, M. (2020). The Current Situation, Problems and Reflection of Online Teaching for Primary and Secondary School Teachers During a Large-scale Epidemic. *China Educational Technology, 5*, 15-21.

Wang, Y., Heffernan, N. (2013). Extending Knowledge Tracing to Allow Partial Credit: Using Continuous versus Binary Nodes (pp. 181-188). In *Proceeding of the 16th Artificial Intelligence in Education*. Berlin: Springer.

Zhong, S. (2020). How Artificial Intelligence Supports Educational Revolution. *China Educational Technology, 3*, 17-24. doi:10.3969/j.issn.1006-9860.2020.03.003.

Yeung, C. K. (2019). Deep-IRT: Make Deep Learning Based Knowledge Tracing Explainable Using Item Response Theory. https://arxiv.org/pdf/1904.11738.pdf.

Yu, S., Wang, H. (2020). How to Better Organize Online Learning in Extreme Situations such as Epidemics. *China Educational Technology, 5*, 6-14. doi:10.3969/j.issn.1006-9860.2020.05.003.

Yudelson, M. V., Koedinger, K. R., & Gordon, G. J. (2013). Individualized Bayesian Knowledge Tracing Models. In *Proceeding of 16th International Conference on Artificial Intelligence in Education* (pp. 171-180). Berlin: Springer.

Zhang, J., Shi, X., King, I., & Yeung, D. (2017). Dynamic Key-Value Memory Networks for Knowledge Tracing. In *Proceedings of the 26th International Conference on World Wide Web* (pp. 765–774). Perth, Australia: International World Wide Web Conferences Steering Committee.