Knowledge or Gaming? Cognitive Modelling Based on Multiple-Attempt Response

Runze Wu
University of Science and Technology of China
Hefei, China
wrz179@mail.ustc.edu.cn

Guandong Xu*
University of Technology, Sydney
Sydney, Australia
Guandong.Xu@uts.edu.au

Enhong Chen†
University of Science and Technology of China
Hefei, China
cheneh@ustc.edu.cn

Qi Liu
University of Science and Technology of China
Hefei, China
qiliuql@ustc.edu.cn

Wan Ng
University of Technology, Sydney
Sydney, Australia
Wan.Ng@uts.edu.au

ABSTRACT

Recent decades have witnessed the rapid growth of intelligent tutoring systems (ITS), in which personalized adaptive techniques are successfully employed to improve the learning of each individual student. However, the problem of using cognitive analysis to distill the knowledge and gaming factor from students learning history is still underexplored. To this end, we propose a *Knowledge Plus Gaming Response Model* (KPGRM) based on multiple-attempt responses. Specifically, we first measure the explicit gaming factor in each multiple-attempt response. Next, we utilize collaborative filtering methods to infer the implicit gaming factor of one-attempt responses. Then we model student learning cognitively by considering both gaming and knowledge factors simultaneously based on a signal detection model. Extensive experiments on two real-world datasets prove that KPGRM can model student learning more effectively as well as obtain a more reasonable analysis.

Keywords

Educational Data Analytics, Intelligent Tutoring Systems, Context-Aware Web-based Learning, Gaming the System, Cognitive Analysis

1. INTRODUCTION

One of the most important innovations in computer aided education during the past decade is intelligent tutoring systems (ITS) [8, 4, 5], which is designed for adaptively providing learners with interactive and customized instruction or feedback. Nowadays, a huge number of ITS, like *Carnegie Learning*[1], *ASSISTments*[2], *Knewton*[3] and *Smart Sparrow*[4], have been built for both novices and experts to learn and self-improve.

A key issue in educational scenarios is to cognitively model student learning from their responses to questions in learning systems, which aims at discovering the knowledge proficiency or learning ability of the students. Recently, one basic assumption about student learning that has been increasingly widely adopted [1, 34] is that: *the response of students in learning systems is synthetically influenced by both knowledge learning, i.e. the proficiency levels of the related knowledge to learn, and gaming strategy, i.e. the ability to use the system itself and solve problems like guessing or retrying until correct.* In other words, each response is assumed to involve one *gaming factor*, i.e. the extent to which one student is "gaming" during his/her response to one question. Studies [7, 19] from pedagogy has revealed the significant impacts of the gaming factor on students' learning performance. Therefore psychometricians developed a series of cognitive models [29, 3, 22] on examination data by considering the gaming factor as a fixed or question-side parameter for modelling student learning. Comparatively, educational data miners employed data mining techniques like feature engineering [37, 2, 14] to detect gaming behaviour in ITS.

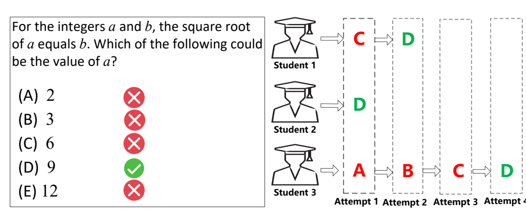

Figure 1: A toy example of student responses in ITS.

*Co-corresponding author.
†Co-corresponding author.

[1]http://www.carnegielearning.com/
[2]https://www.assistments.org/
[3]https://www.knewton.com/
[4]https://www.smartsparrow.com/

WWW'17 Companion, April 3–7, 2017, Perth, Australia.
ACM 978-1-4503-4914-7/17/04.
http://dx.doi.org/10.1145/3041021.3054156

Despite the importance of the previous studies, there are some existing limitations. Most traditional cognitive models mainly focus on one-attempt response data, e.g. examination, which is quite different from the multiple-attempt response data encountered in the ITS context. Taking the toy example shown in Fig. 1, it can be seen that some students answer correctly on the first attempt (e.g. Student 2), which forms *one-attempt responses* (OAR); the others who fail on the first attempt, keep trying until correct (e.g. Student 3), forming *multiple-attempt responses* (MAR). Apparently utilising the first-attempt or one-attempt responses, the current psychometrical models are unable to capture a full view of the gaming factor for more precise analysis (e.g., it is hard to distinguish whether Student 2 is "gaming"). In contrast, MAR, which explicitly conveys the attempt details, can be analysed to obtain another view of the gaming factor (e.g., Student 3 is probably "gaming" by trying each of the possible answers). Moreover most educational data miners treat gaming behaviour in ITS as a classification task based on feature engineering, instead of cognitively modelling the students to discover their knowledge proficiency or learning ability. Thus, it is of significant importance to capture the full view of the gaming factors from MAR and then incorporate into the whole cognitive modelling process. To this end, there are several challenges: 1) how to measure the explicit gaming factor from MAR; 2) based on 1), how to further infer the implicit gaming factor from the existing OAR; and 3) how to cognitively model student learning by incorporating knowledge with the obtained gaming factors?

To address these challenges, we propose a *Knowledge Plus Gaming Response Model* (KPGRM) based on MAR to model student learning cognitively. Based on educational domain knowledge on the gaming behaviour, we adopt a P-value evidence based method to measure the gaming factor using four observable aspects and then aggregate them as the explicit gaming factor of MAR. Then we employ collaborative filtering techniques to indirectly infer the implicit gaming factor of OAR. Furthermore, a simple signal detection model is utilized to cognitively fuse both the knowledge and gaming impacts on student learning. Model parameters are estimated by a Markov Chain Monte Carlo (MCMC) means. The main contributions of this paper are as follows:

- To the best of our knowledge, this is the first comprehensive attempt at discovering implicit and explicit gaming factors and combining knowledge and gaming for student learning modelling to obtain more precise and reasonable cognitive analysis.
- We propose a cognitive model KPGRM, which employs educational domain knowledge and collaborative filtering for evidentially extracting the gaming factor from MAR and OAR, and links students' responses to knowledge proficiency based on a simple signal detection model.
- We design an effective MCMC sampling algorithm for parameter estimation and conduct extensive experiments on real-world datasets to verify the effectiveness of KPGRM.
- We analyse the reasonability of the extracted gaming factor as well as study the knowledge and gaming impacts on question difficulty based on KPGRM.

Overview. The rest of this paper is organized as follows. In Section 2, we introduce the related work on student learning modelling and gaming factor. In Section 3, we formally define our targeted issue. Section 4 details the whole framework of our KPGRM. Section 5 shows the experimental results to verify the effectiveness and reasonability of our approach. Conclusions are given in Section 6.

2. RELATED WORK

We introduce the existing related work from two aspects: student learning modelling and the gaming factor in ITS.

2.1 Student Learning Modelling

In educational psychology, many psychometrical models [13, 35] have been developed to mine students' knowledge proficiency level from responses to questions. These models can be roughly divided into two categories: continuous ones and discrete ones. The fundamental continuous models are *item response theory* (IRT) models [29, 3, 15], which characterize students by a continuous variable, i.e. knowledge ability, and use a logistic function to model the probability that a student will correctly solve a problem. For the discrete models, the basic method is *deterministic inputs, noisy "and" gate model* (DINA) [17, 22]. DINA describes a student by a latent binary vector which denotes whether (s)he has mastered the skills required by the problem with given prior information. In addition, some general approaches are proposed for either fusing the continuous and the discrete models [12] or incorporating more complex questions like free-response ones [39]. In ITS context, [9, 10] proposed *Bayesian knowledge tracing* (BKT) models based on hidden Markov models and [6] designed a variant of IRT model, *learning factor analysis* (LFA).

However, most of the current psychometrical models consider only the first-attempt responses and simply ignore the subsequent multiple-attempt ones hence, as shown in Fig. 1, valuable information is not fully exploited. In this paper we take into account multiple-attempt data to extract the gaming factor into modelling student learning.

2.2 Gaming Factor

Gaming-the-system or the gaming factor, which harms the effectiveness of learning systems to some extent, universally exists and also draw a lot of attention from educational and data mining fields. [7, 19] studied impacts of the gaming factor on students' learning performance via real-world experiments with pretests and posttests. Traditional psychometrical models [13] usually regard the gaming factor as guessing which is estimated by fixing a heuristic value (e.g. 1/#*option*) or parameterizing it from question-side information (e.g. 3PL-IRT [3]). With the richer features in ITS (e.g. activity logs), educational data miners have adopted some feature engineering methods [37, 2, 14] to detect gaming factor. Further, [20] takes into account response time and models student learning with the hidden motivation.

Nevertheless, the existing approaches to handle the gaming factor either simply view it as a detection task solved by classifiers, or model student learning with additional limited information from only the first-attempt responses. Here in this work we extract the gaming factor by utilising multiple-attempt data and then incorporate into the whole cognitive modelling process.

3. PROBLEM FORMULATION

Suppose we have M students who answer N questions[5] in an ITS and R stands for all the responses where R_{ji} denotes Student j's response to Question i. Here, R_{ji} is made up of an ordered sequence of tuples defined as:

$$R_{ji} = \{\langle c_{jik}, r_{jik}, t_{jik} \rangle | k \in \mathbb{Z}, 1 \leq k \leq K_{ji}\},$$

where c_{jik}, r_{jik} and t_{jik} represent the actual content, the auto-generated label (1 or 0) indicating correct or not and the time stamp of the kth attempt of Student j to respond to Question i, respectively. $K_{ji} \in \mathbb{Z}^+$ denotes the length of the response sequence of Student j on Question i. The key challenge and goal of our formulation is to effectively measure or infer the gaming factor of each response and model students' knowledge structure then precisely estimate the proficiency level of each student, where $\theta_j \in \mathbb{R}$ represents the knowledge ability of Student j.

Table 1: Some important notations.

Notation	Description
R_{ji}	The response of Student j to Question i
$\mathbf{R}^*$	The set of all the MAR
r_{ji1}	The label of the first-attempt response of Student j to Question i
θ_j, ϑ_j	The knowledge and gaming ability of Student j
β_i, γ_i	The knowledge and gaming difficulty of Question i
g_{ji}	The gaming factor of Response R_{ji}

4. KNOWLEDGE PLUS GAMING RESPONSE MODEL

To better model students' knowledge structure and estimate their proficiency level with the multiple-attempt responses (MAR), in this section, we introduce the detail of our proposed KPGRM method. Fig. 2 shows the whole schema of our model framework. To be specific, we extract the gaming factor firstly from multiple-attempt (explicit) and one-attempt (implicit) responses and we model student learning by combining the knowledge and gaming factors. Tab. 1 lists some important notations and each step of KPGRM is elaborated in the following subsections.

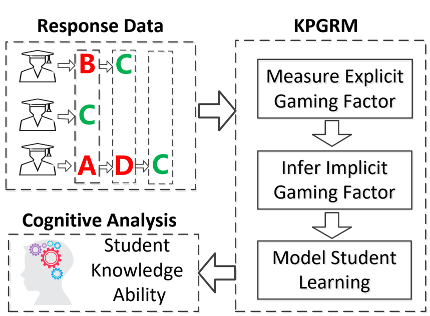

Figure 2: The framework of KPGRM.

4.1 Explicit Gaming Factor of MAR

Different from most traditional cognitive models which focus on only first-attempt responses, in this subsection, we take MAR into account to extract the gaming factor.

[5]To simplify, we only focus on problems of multiple-choice questions involving one general topic.

Here we give a formal definition to measure the explicit gaming factor from each MAR. According to the existing literature [11, 2], gaming behaviour could be represented by *keep answering*, *systematically* and *quickly* until an identified correct response allows the student to move to the next question. Based on the existing domain knowledge of gaming and the availability of current data, we mainly focus on four characteristics and assumptions when measuring the gaming factor from MAR: 1) the more attempts in one response, the higher the gaming factor of the relevant student ("keep answering"); 2) the less time taken to answer, the higher the gaming factor ("quickly"); 3) the more transitions in one response, the higher the gaming factor ("systematically"); 4) the higher the coverage of the given options, the higher the gaming factor ("systematically"). Thus we could define the explicit gaming factor for each MAR as an aggregation function of these four characteristics. Formally, given a MAR R_{ji} of Student j to Question i, we define the gaming factor from $R_{ji} \in \mathbf{R}^*$ as

$$g_{ji} = F(Len(R_{ji}), Spd(R_{ji}), Trs(R_{ji}), Cov(R_{ji})), \quad (1)$$

where g_{ji} denotes the gaming factor of the response of Student j to Question i and $F(\cdot)$ is an aggregation function. Note that $\mathbf{R}^* = \{R_{ji} | K_{ji} \geq 2\}$ is a set of all the MAR of all the students and questions. Here $Len(R_{ji})$,$Spd(R_{ji})$,$Trs(R_{ji})$ and $Cov(R_{ji})$ represent four characteristics from MAR R_{ji}, i.e. length of attempt sequence, speed of answering, transition and coverage of all given options through the whole sequence. For convenience of normalization and calculation, we adopt a statistical P-value based approach [40] to describe the four aspects of each MAR which is specified in detail as follows.

LENGTH. As mentioned previously, the larger length of MAR represents a higher gaming factor. Here K_{ji}, i.e. the length of the multiple-attempt response of Student j to Question i, is assumed to follow the Poisson distribution, $K_{ji} \sim \mathcal{P}(\lambda_K)$, where λ_K can be learned by the maximum-likelihood estimation (MLE) method from the observations in the given records. Then, we can define gaming evidence of length by

$$Len(R_{ji}) = 1 - P(\mathcal{P}(\lambda_K) \geq K_{ji}), \quad (2)$$

where we can obtain the P-value by calculating $P(\mathcal{P}(\lambda_K) \geq K_{ji})$. Accordingly, a smaller P-value involving a longer attempt sequence means a higher gaming factor.

SPEED. As discussed previously, faster answering signifies a higher gaming factor. Here we use the average time of each attempt to capture the SPEED characteristics by

$$\bar{ts}_{ji} = \frac{t_{jiK_{ji}} - t_{ji1}}{K_{ji} - 1}. \quad (3)$$

where $t_{jiK_{ji}}$ and t_{ji1} are the time stamps of the last and the first attempt. Here we assume that $\bar{ts}_{ji}$, i.e. the average time of MAR of Student j to Question i, follows the Gaussian distribution, $\bar{ts}_{ji} \sim \mathcal{N}(\mu_{ts}, \sigma_{ts}^2)$, where the parameter μ_{ts} and σ_{ts}^2 can be learned by the MLE method from the observations of $\bar{ts}_{ji}$ in the given records. Then, we can define gaming evidence of speed by

$$Spd(R_{ji}) = 1 - P(\mathcal{N}(\mu_{ts}, \sigma_{ts}^2) \leq \bar{ts}_{ji}). \quad (4)$$

where we obtain the P-value by calculating $P(\mathcal{N}(\mu_{ts}, \sigma_{ts}^2) \leq \bar{ts}_{ji})$. Similarly, a multiple-attempt response with a smaller value of this P-value has a higher gaming factor.

TRANSITION. As assumed previously, the more transitions mean a higher gaming factor. Here we assume that the transition tr_{ji}, i.e. the number of changes between attempts of MAR of Student j to Question i, follows the Poisson distribution, $tr_{ji} \sim \mathcal{P}(\lambda_{tr})$, where the parameter λ_{tr} and can be learned by the MLE method from the observations of tr_{ji} in the given records. Then, we can define gaming evidence of transition by

$$Trs(R_{ji}) = 1 - P(\mathcal{P}(\lambda_{tr}) \geq tr_{ji}). \quad (5)$$

where we obtain the P-value by calculating $P(\mathcal{P}(\lambda_{tr}) \geq tr_{ji})$. Similarly, a multiple-attempt response with a smaller value of this P-value has a higher gaming factor.

COVERAGE. As assumed previously, higher coverage of all given options implies a higher gaming factor. Here we assume that the coverage cov_{ji}, i.e. the percentage of options of MAR that Student j has tried to respond to Question i, follows the Gaussian distribution, $cov_{ji} \sim \mathcal{N}(\mu_{cov}, \sigma^2_{cov})$, where the parameter μ_{cov} and σ^2_{cov} and can be learned by the MLE method from the observations of cov_{ji} in the given records. Then, we can define gaming evidence of coverage by

$$Cov(R_{ji}) = 1 - P(\mathcal{N}(\mu_{cov}, \sigma^2_{cov}) \geq cov_{ji}). \quad (6)$$

where we obtain the P-value by calculating $P(\mathcal{N}(\mu_{cov}, \sigma^2_{cov}) \geq cov_{ji})$. Similarly, a multiple-attempt response with a smaller value of this P-value has a higher gaming factor.

After extracting the four aspects of evidence of the gaming factor, the next challenge is how to combine them, i.e. to figure out a proper function $F(\cdot)$. In fact, there are many supervised evidence aggregation methods in the literature [36, 16] which depend on labelled training data. For convenience, instead, we adopt an unsupervised aggregation approach based on the similarity between the extracted evidence.

Specifically, we choose a linear combination of all the evidences of MAR of Student j to Question i as the aggregation function $F(\cdot)$ as follows:

$$F(\Phi_{ji}(1), \Phi_{ji}(2), \cdots, \Phi_{ji}(E)) = \sum_{e=1}^{E} \omega_e \Phi_{ji}(e), \text{ s.t. } \sum_{e=1}^{E} \omega_e = 1, \quad (7)$$

where $\Phi_{ji}(e)$ denotes the eth evidence and $\omega_e \in [0, 1]$ is the corresponding weight. Note that in our case $E = 4$ for our four defined pieces of evidence. Next we introduce our unsupervised method to learn the proper $\{\omega_e\}$.

Here, we adopt an intuitive assumption as *Consistent Better*, which has been proved effective in many applications, for our evidence aggregation. To be specific, we assume that *effective evidence should have a similar evidence score for each MAR, while poor evidence will produce different scores.* Therefore, evidence that tends to be consistent with the majority of evidence will be assigned higher weights and evidence that tends to disagree will be assigned lower weights. Then we can measure the consistence of each evidence $\Phi_{ji}(e)$ using the variance-like measure

$$\Delta_{ji}(e) = (\Phi_{ji}(e) - \bar{\Phi}_{ji})^2, \quad (8)$$

where $\bar{\Phi}_{ji}$ is the average score of all the defined types of evidence. In line with *Consistent Better*, $\Phi_{ji}(e)$ should be given a larger weight if $\Delta_{ji}(e)$ is small. Thus, we can redefine the evidence aggregation problem as an optimization problem that minimizes the weighted variance of the evidence over all the MAR

$$\arg\min_{\omega} \sum_{R_{ji} \in \mathbf{R}^*} \sum_{e=1}^{E} \omega_e \Delta_{ji}(e), \quad (9)$$
$$\text{s.t. } \sum_{e=1}^{E} \omega_e = 1; \forall \omega_e \in [0, 1].$$

Here, we employ a popular gradient based approach [23, 24] with exponentiated updating to solve this problem.

4.2 Implicit Gaming Factor of OAR

With the explicit gaming factor measured from MAR, in this subsection, we specify how to infer the implicit gaming factor of OAR for our KPGRM model.

Different from MAR with richer information, it is hard to distinguish the "gaming" OAR, which is usually implicit (e.g. a student may answer correctly by guessing on the first attempt). Here we determine the inference of the gaming factor of OAR by collaborative filtering (CF).

CF assumes that each user and each item are all related so that similar users have similar preferences while a user will likely like items that are similar to the currently preferred ones. In recommender systems, the key bridge connecting users and items is user-item interaction like consuming, which can be utilized to model preferences by CF [30]. Similarly, we could regard each student and question as a user and item then the gaming factor represents the interaction between users and items. Thus we redefine the gaming factor inference problem as the interaction prediction problem. There are lots of existing predictive methods like neighbourhood-based [38] and user/item-based CF [27]. However, these memory-based methods are more suitable for the top-N recommendation problem. For our case we adopt the latent factor model [28, 25] for our inference task due to the powerful prediction ability of this model-based method.

Specifically, we map each student and question into a new d-dimension space which depicts the latent psychological characteristics of students in the learning process and the corresponding latent properties of the questions. Formally, we use $U \in \mathbb{R}^{m \times d}$ and $V \in \mathbb{R}^{n \times d}$ to represent each student and question in the latent space, with column vectors U_j and V_i denoting latent feature vectors of Student j and Question i, respectively. A probabilistic linear model with Gaussian observation noise is adopted to define the conditional distribution over the explicit gaming factors as

$$P(G|U, V, \sigma^2_g) = \prod_{R_{ji} \in \mathbf{R}^*} \mathcal{N}(g_{ji}|U_j V_i^T, \sigma^2_g), \quad (10)$$

where $G \in \mathbb{R}^{m \times n}$ is the gaming factor matrix consisting of the gaming factors of each student on each question and σ^2_g is the variance of the gaming factors.

Next we maximize the logarithm of the posterior likelihood over observations (i.e. the explicit gaming factor previously measured from MAR in Eq. (1)) by minimizing the following objective function to estimate U and V.

$$E = -\frac{1}{2} \sum_{R_{ji} \in \mathbf{R}^*} (g_{ji} - U_j V_i^T)^2 + \frac{\lambda_U}{2} \sum_{j=1}^{M} \|U_j\|^2 + \frac{\lambda_V}{2} \sum_{i=1}^{N} \|V_i\|^2, \quad (11)$$

where λ_U and λ_V are the regularization parameters. We adopt a stochastic gradient descent in U and V for optimization. Then the implicit gaming factor of all OAR can easily be inferred by learnt U and V.

To this point, we have proposed our method to extract the gaming factor from all the responses, either MAR or OAR, by a direct measure or an indirect inference. We can summarize it as the following equation:

$$g_{ji} = \begin{cases} F(Len(R_{ji}), Spd(R_{ji}), Trs(R_{ji}), Cov(R_{ji})) & \text{if } R_{ji} \in \mathbf{R}^*; \\ U_j V_i^T & \text{if } R_{ji} \notin \mathbf{R}^*. \end{cases} \quad (12)$$

4.3 Model Student Learning

With the gaming factor extracted from all the response, in this subsection, we incorporate the gaming factor into student learning modelling in a more reasonable way.

As discussed in the introduction, students can answer a question by either using genuinely learned knowledge or by simply gaming the system. However, most of the traditional learning models, which neglect attempts subsequent to the first one, do not capture the gaming factor for a more accurate estimate of students' knowledge ability. To mitigate this issue, we first extract the gaming factor of all the responses then model both knowledge and gaming ability simultaneously.

Specifically, inspired by much of the existing work from education and psychology [32, 22, 31], we adopt a simple *signal detection model* for our task to "detect" gaming factor g_{ji} from noisy observations, i.e. first-attempt response r_{ji1}. Then let us consider two extreme conditions:

1) the student answers the question correctly without any gaming factor, where we model student learning as follows:

$$\eta_{ji} \overset{\text{def}}{=} P(r_{ji1} = 1 | g_{ji} = 0) = \frac{1}{1 + \exp\{-(\theta_j - \beta_i)\}}. \tag{13}$$

Here we adopt a simple one-parameter logistic IRT (1PL-IRT) model in which θ_j and β_i represent the knowledge ability of Student j and the relevant difficulty of Question i, respectively. To be specific, remembering, understanding or mastering some knowledge topics (e.g. vocabulary and concepts) is a knowledge ability while the knowledge difficulty of one question depends on the related knowledge topics. Note that we choose η_{ji} to denote the probability.

2) the student answers correctly and completely by gaming, where we model student learning as follows:

$$\zeta_{ji} \overset{\text{def}}{=} P(r_{ji1} = 1 | g_{ji} = 1) = \frac{1}{1 + \exp\{-(\vartheta_j - \gamma_i)\}}. \tag{14}$$

Similarly, we also choose a 1PL-IRT model in which ϑ_j and γ_i represent the gaming ability of Student j and the relevant difficulty of Question i, respectively. Specifically, how to pick or guess the right answer quickly is a gaming ability while the gaming difficulty of one question usually is based on the question design including structure, description and option settings. Note that we also choose ζ_{ji} to denote the probability.

Then, assuming the statistical independence of responses on each question conditioned on the students' ability [29, 22, 13], we employ Bernoulli distribution to model all the first-attempt responses, which are either right or wrong, as follows:

$$P(r_{ji1} = 1 | \theta_j, \beta_i, \vartheta_j, \gamma_i, g_{ji}) = \eta_{ji}^{1-g_{ji}} \zeta_{ji}^{g_{ji}}, \tag{15}$$

where η_{ji} and ζ_{ji} stand for the probability that Student j responds to Question i correctly based on knowledge learning and gaming strategy, respectively. The model will degenerate as an ordinary IRT model using traditional settings without the consideration of gaming if $g_{ji} = 0$. Meanwhile, Eq. 15 can also be viewed as a variant of the non-compensatory bi-dimensional item response model [33] where gaming ability also serves as a kind of latent trait.

Summary. We first measure the explicit gaming factor from MAR and then infer the implicit gaming factor from OAR. Next based on a simple signal detection model, we fuse both the knowledge and gaming to model student learning. As shown in Fig. 3, what we can observe from Student j and Question i is the response R_{ji}, where we obtain the first-attempt response r_{ji1} and the extracted gaming factor g_{ji}. In this paper we model student learning from two aspects: knowledge, i.e. ability θ_j and the relevant difficulty β_i, and gaming, i.e. ability ϑ_j and the corresponding difficulty γ_i. As proposed previously, we assume that the response of each student to each question is affected by genuine knowledge learning η_{ji} and artful gaming strategy ζ_{ji}.

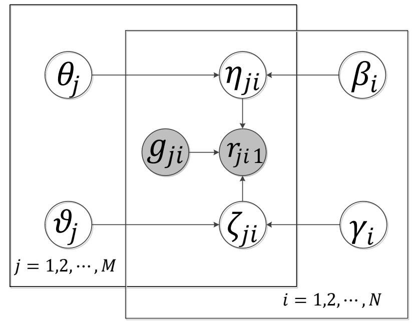

Figure 3: The graphic model of KPGRM.

4.4 Model Estimation

In this subsection, we introduce an effective training algorithm using MCMC for the proposed KPGRM model, that is, to estimate the unshaded variables in Fig. 3. Specifically, we assume the prior distributions of the parameters in KPGRM as follows:

$$\begin{aligned} \theta_j &\sim \mathcal{N}(\mu_\theta, \sigma_\theta^2), \beta_i \sim \mathcal{N}(\mu_\beta, \sigma_\beta^2); \\ \vartheta_j &\sim \mathcal{N}(\mu_\vartheta, \sigma_\vartheta^2), \gamma_i \sim \mathcal{N}(\mu_\gamma, \sigma_\gamma^2). \end{aligned} \tag{16}$$

The functional forms of the prior distributions are chosen for convenience, and the associated hyperparameters are selected to be reasonably vague within the range of the realistic parameters. Then, the joint posterior distribution of θ,ϑ,β and γ given the responses R is as follows:

$$P(\theta, \vartheta, \beta, \gamma | R) \propto L(\theta, \vartheta, \beta, \gamma) P(\theta) P(\vartheta) P(\beta) P(\gamma). \tag{17}$$

where L is the joint likelihood function of KPGRM which, according to Eq. 15, is defined as follows:

$$L(\theta, \vartheta, \beta, \gamma) = \prod_{j,i} (\eta_{ji}^{1-g_{ji}} \zeta_{ji}^{g_{ji}})^{r_{ji1}} (1 - \eta_{ji}^{1-g_{ji}} \zeta_{ji}^{g_{ji}})^{1-r_{ji1}}. \tag{18}$$

The full conditional distributions of the parameters given the observations and the rest of parameters are as follows:

$$P(\theta | R, \vartheta, \beta, \gamma) \propto L(\theta, \vartheta, \beta, \gamma) P(\theta), \tag{19}$$
$$P(\vartheta | R, \theta, \beta, \gamma) \propto L(\theta, \vartheta, \beta, \gamma) P(\vartheta), \tag{20}$$
$$P(\beta | R, \theta, \vartheta, \gamma) \propto L(\theta, \vartheta, \beta, \gamma) P(\beta), \tag{21}$$
$$P(\gamma | R, \theta, \vartheta, \beta) \propto L(\theta, \vartheta, \beta, \gamma) P(\gamma). \tag{22}$$

Finally, we propose a Metropolis-Hastings (M-H) based MCMC algorithm [18] for parameter estimation by Alg. 1. To be specific, we first randomize all the parameters as the initial values. Then, using observed responses R, we compute the full conditional probability of knowledge ability θ, the relevant difficulty β, the gaming ability ϑ and the corresponding difficulty γ. Next, the acceptance probability of the samples can also be calculated based on the M-H algorithm. In this way, we estimate the parameters with the MCMC formed through sampling.

Table 2: Datasets Summary.

Dataset	# Student	# Question	Avg #Attempt	Avg Response Time (sec)	Avg #Transition	Avg Coverage
Alone	140	16	2.44	35.03	1.89	0.43
Earth	164	14	2.31	39.56	1.38	0.60

Algorithm 1 Sampling algorithm for KPGRM.

Input: all the response R and gaming factor G
Output: samples of $\theta, \vartheta, \beta, \gamma$
Initialize $\theta_0, \vartheta_0, \beta_0, \gamma_0$ with random values
for $t = 1, 2, \cdots, T$ **do**
  Draw $\theta_t \sim U(\theta_{t-1} - \delta_\theta, \theta_{t-1} + \delta_\theta)$, and accept θ_t with the probability:
  $min\{1, \frac{L(\theta_t,\vartheta_{t-1},\beta_{t-1},\gamma_{t-1})P(\theta_t)}{L(\theta_{t-1},\vartheta_{t-1},\beta_{t-1},\gamma_{t-1})P(\theta_{t-1})}\}$.
  Draw $\vartheta_t \sim U(\vartheta_{t-1} - \delta_\vartheta, \vartheta_{t-1} + \delta_\vartheta)$, and accept ϑ_t with the probability:
  $min\{1, \frac{L(\theta_{t-1},\vartheta_t,\beta_{t-1},\gamma_{t-1})P(\vartheta_t)}{L(\theta_{t-1},\vartheta_{t-1},\beta_{t-1},\gamma_{t-1})P(\vartheta_{t-1})}\}$.
  Draw $\beta_t \sim U(\beta_{t-1} - \delta_\beta, \beta_{t-1} + \delta_\beta)$, and accept β_t with the probability:
  $min\{1, \frac{L(\theta_{t-1},\vartheta_{t-1},\beta_t,\gamma_{t-1})P(\beta_t)}{L(\theta_{t-1},\vartheta_{t-1},\beta_{t-1},\gamma_{t-1})P(\beta_{t-1})}\}$.
  Draw $\gamma_t \sim U(\gamma_{t-1} - \delta_\gamma, \gamma_{t-1} + \delta_\gamma)$, and accept γ_t with the probability:
  $min\{1, \frac{L(\theta_{t-1},\vartheta_{t-1},\beta_{t-1},\gamma_t)P(\gamma_t)}{L(\theta_{t-1},\vartheta_{t-1},\beta_{t-1},\gamma_{t-1})P(\gamma_{t-1})}\}$.
  if convergence criterion meets **then**
    return
  end if
end for
return

5. EXPERIMENT

We first prove the effectiveness of KPGRM against the baseline approaches by predicting student performance; then, we further conduct gaming factor and question difficulty analysis to demonstrate the reliability of our method.

5.1 Setup

The real-world MAR datasets in our experiment are collected from *Smart Sparrow*[6] where students enrolled in different schools study two science courses *Are we alone* and *Earth*. To alleviate sparsity, we construct the datasets by filtering relatively inactive students and questions. Then we denote the two obtained datasets as *Alone* and *Earth*. Each of the datasets contains the actual response content, the label indicating correct or not and the time stamp of each student to each question at each attempt. A brief summary of each dataset is shown in Tab. 2. And Fig. 4 shows an overview of the two datasets, where each subfigure is a matrix depicting the number of attempts of each response, each row denotes a student and each column represents a question. The yellower one means more attempts of one response while the bluer one indicates less attempts.

For the prior distributions of the parameters in Alg. 1, we set the hyperparameters as follows:

$$\mu_\theta = 0, \sigma_\theta^2 = 1; \mu_\beta = 0, \sigma_\beta^2 = 2;$$
$$\mu_\vartheta = 0, \sigma_\vartheta^2 = 1; \mu_\gamma = 0, \sigma_\gamma^2 = 2.$$

In these experiments, we set the number of iterations to 5,000 and estimate the parameters based on the last 2,500 samples to guarantee the convergency of the Markov chain.

[6]https://www.smartsparrow.com/

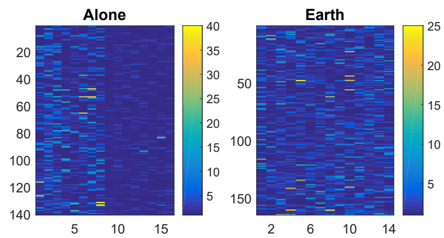

Figure 4: Overview of the datasets.

5.2 Model Evaluation

To evaluate the performance of our KPGRM in terms of cognitive modelling, we choose *Predicting Student Performance*, one of the key tasks in educational systems [21, 6], compared with some popular methods from psychometrics and data mining as baseline methods. We adopt three metrics from different perspectives: *root mean square error* (RMSE), *classification accuracy* (ACC) and *area under an ROC curve* (AUC).

Specifically, we employ 5-fold cross validation on each of the datasets where one of five folds is targeted for testing and the remaining parts for training in each pass. The baseline methods are as follows:

- *IRT*: [29, 3] a cognitive diagnosis method modelling students' latent traits and the parameters of questions such as difficulty.
- *PMF*: [28] probabilistic matrix factorization is a latent factor model projecting students and questions into a low-dimensional space.
- *NMF*: [26] non-negative matrix factorization is a latent non-negative factor model and can be viewed as a topic model.
- *LFA*: [6] an educational data mining model considering the different impacts of the defined knowledge factors on student performance.

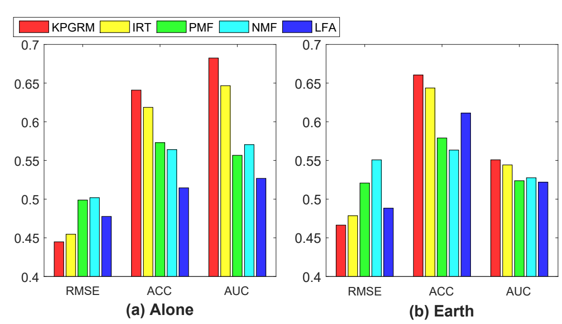

Figure 5: The comparison of prediction performance.

For the purpose of comparison, we record the best performance of each algorithm by tuning their parameters and Fig. 5 shows the prediction results of our KPGRM and baseline approaches on the two datasets. We observe that, over all the datasets, KPGRM performs the best. Specifically, when considering cognitive assumptions ("knowledge plus gaming") it outperforms PMF and NMF, and when incorporating the gaming factor, it outperforms IRT and LFA. Of the baseline methods, IRT, as the classical psychometrical model, outperforms the others while LFA obtains a relatively poorer result by modelling only one general knowledge factor. In summary, considering the gaming factor, our KPGRM captures the characteristics of students more precisely and it is also more suitable for real-world scenarios.

5.3 Gaming Factor Analysis

In addition to model evaluation, we conduct a cognitive analysis on the gaming factor. Firstly, we check the effectiveness of our method to extract the gaming factor. Due to the lack of ground truth, we adopt human coding [1] for indirect verification. Specifically, we randomly choose 20 MAR from *Alone*, and ask 11 volunteers (educational researchers and graduate students) to scrutinize the detail of each attempt of an MAR and allocate a gaming score (where 1, 0.5 and 0 denote "gaming", "not sure" and "no gaming", respectively). Achieving acceptable inter-rater reliability (Fleiss' $\kappa = 0.73$), we compute the average AUC by considering MAR with a gaming score of 1, 0.5 and 0 as positive, neutral and negative case. Tab. 3 shows the results computed by $Len(\cdot), Spd(\cdot), Trs(\cdot), Cov(\cdot)$ and our aggregated measure $F(\cdot)$. From this comparison we can observe that our aggregated measure for the gaming factor $F(\cdot)$ is the most consistent with human coding.

Table 3: The comparison with human coding.

Measure	$Len(\cdot)$	$Spd(\cdot)$	$Trs(\cdot)$	$Cov(\cdot)$	$F(\cdot)$
avgAUC	0.958	0.723	0.932	0.945	**0.966**

(Fleiss' $\kappa = 0.73$)

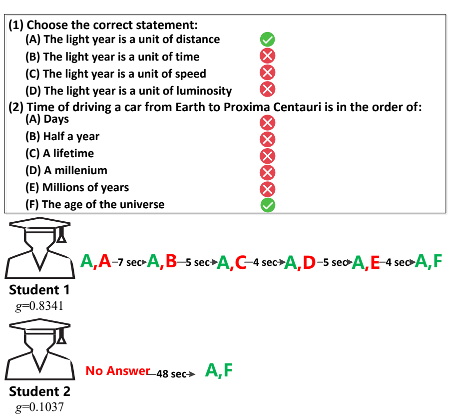

Figure 6: Two cases with different gaming factors.

Furthermore, we also study two cases with different gaming factors. As shown in Fig. 6, the question above comprises two steps which contain four and six options, respectively. The chosen option of each attempt and time spent (seconds) between each attempt of two students are also presented below. We can observe that Student 1 tries each of the given options of Step 2 systematically and quickly while being very sure of the correct answer of Step 1, hence the extracted gaming factor is 0.8341. On the contrary, Student 2 forgets to input the answer first and then spends a relatively longer time figuring out the correct options at the second attempt, hence the gaming factor is much lower at 0.1037. From the comparison of the two real-world cases we can see that the extracted gaming factor is very intuitive: the more significant gaming behaviour, the larger gaming factor.

5.4 Question Difficulty Analysis

Based on our KPGRM framework, we also analyse the question difficulty by considering gaming impacts. As discussed in Section 4.3, question difficulty comes from two aspects: knowledge learning and gaming strategy. It is of significant importance for learning systems to delicately design questions for eliminating gaming impacts and capturing the actual level of students for better personalized instruction. Fig. 7 shows the relationship between the two question properties, i.e. the number of steps and options[7], and the two kinds of difficulties from *Alone*.

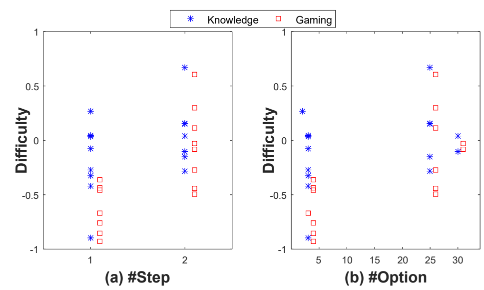

Figure 7: The relationship between two kind of difficulties and two question properties.

We can observe that the gaming difficulty of the questions with less steps or options is more likely to be lower, which means it is easier for students to solve by using an artful gaming strategy such as guessing. Pearson's ρ between the gaming difficulty and the number of steps and options of the questions is 0.7375 and 0.7287, respectively. On the other hand, the knowledge difficulty of the questions is not significantly related to the question properties. Pearson's ρ between the gaming difficulty and the number of steps and options of the questions is 0.3626 and 0.3254, respectively. The results conform to the intuition: the more complicated the design (including more steps or options), the higher the gaming difficulty of the question. Thus our KPGRM can also be utilized to target the questions with low gaming difficulty and improve the system design to enhance the effectiveness of the ITS.

5.5 Discussion

Note that the generic idea of our work is to build a cognitive model to discover the actual learning ability of students by distinguishing the effects of different factors, i.e. knowledge and gaming in the current scenario. Apparently student learning activity as a sophisticated cognitive process, involves a lot of psychological factors, which however, is out of the focus of this work. In practice, the outcome of our

[7]Multiply the number of options of each step if the question has multiple steps.

model can be applied beyond the cognitive modelling itself, for example, for evaluating the ITS content design.

On the other hand, there is still room for improvement. Our KPGRM only considers four aspects of MAR to measure the gaming factor, so we will try to utilise more information to enhance the measurement. In addition, our KPGRM computes the gaming factor directly or indirectly and regards it as observed, and we will build a more robust model by modelling the gaming factor as partially observed. Furthermore, many more factors impacting student response are underexplored beyond knowledge and gaming.

6. CONCLUSION

In this paper, we designed a Knowledge Plus Gaming Response Model, KPGRM, to precisely explore the gaming factor in student learning based on MAR data. Specifically, we first measured the explicit gaming factor from MAR by an aggregated P-value based method and inferred the implicit gaming factor from OAR. Next, combining the extracted gaming factor, we constructed a novel signal detection response model to precisely describe student learning. Finally we conducted extensive experiments to prove the effectiveness of our method, cognitively analysed the gaming factor and studied the gaming difficulty of the questions. We expect this work could lead to more future studies.

Acknowledgements

This research was partially supported by grants from the ARC Linkage Project (Grant No. LP140100937), the National Science Foundation for Distinguished Young Scholars of China (Grant No. 61325010), and the National Natural Science Foundation of China (Grant No. U1605251 and 61672483).

7. REFERENCES

[1] R. S. Baker, A. T. Corbett, and K. R. Koedinger. Detecting student misuse of intelligent tutoring systems. In *International Conference on Intelligent Tutoring Systems*, pages 531–540. Springer, 2004.

[2] R. S. Baker, A. T. Corbett, K. R. Koedinger, and A. Z. Wagner. Off-task behavior in the cognitive tutor classroom: when students game the system. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, 383–390. ACM, 2004.

[3] A. Birnbaum. Some latent trait models and their use in inferring an examinee's ability. *Statistical theories of mental test scores*, 1968.

[4] P. Brusilovsky, E. Schwarz, and G. Weber. Elm-art: An intelligent tutoring system on world wide web. In *International Conference on Intelligent Tutoring Systems*, pages 261–269. Springer, 1996.

[5] H. Burns, C. A. Luckhardt, J. W. Parlett, and C. L. Redfield. *Intelligent tutoring systems: Evolutions in design*. Psychology Press, 2014.

[6] H. Cen, K. Koedinger, and B. Junker. Learning factors analysis–a general method for cognitive model evaluation and improvement. In *International Conference on Intelligent Tutoring Systems*, pages 164–175. Springer, 2006.

[7] C.-H. Chen, G.-Z. Liu, and G.-J. Hwang. Interaction between gaming and multistage guiding strategies on students' field trip mobile learning performance and motivation. *British Journal of Educational Technology*, 2015.

[8] W. J. Clancey. Methodology for building an intelligent tutoring system. *Methods and tactics in cognitive science*, pages 51–84, 1984.

[9] A. T. Corbett and J. R. Anderson. Knowledge tracing: Modeling the acquisition of procedural knowledge. *User modeling and user-adapted interaction*, 4(4):253–278, 1994.

[10] R. S. d Baker, A. T. Corbett, and V. Aleven. More accurate student modeling through contextual estimation of slip and guess probabilities in bayesian knowledge tracing. In *International Conference on Intelligent Tutoring Systems*, pages 406–415. Springer, 2008.

[11] R. S. d Baker, A. T. Corbett, K. R. Koedinger, S. Evenson, I. Roll, A. Z. Wagner, M. Naim, J. Raspat, D. J. Baker, and J. E. Beck. Adapting to when students game an intelligent tutoring system. In *International Conference on Intelligent Tutoring Systems*, pages 392–401. Springer, 2006.

[12] J. De La Torre. The generalized dina model framework. *Psychometrika*, 76(2):179–199, 2011.

[13] L. V. DiBello, L. A. Roussos, and W. Stout. 31a review of cognitively diagnostic assessment and a summary of psychometric models. *Handbook of statistics*, 26:979–1030, 2006.

[14] J. S. Downs, M. B. Holbrook, S. Sheng, and L. F. Cranor. Are your participants gaming the system?: screening mechanical turk workers. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 2399–2402. ACM, 2010.

[15] S. E. Embretson and S. P. Reise. *Item response theory for psychologists*. Psychology Press, 2013.

[16] Y. Ge, H. Xiong, C. Liu, and Z.-H. Zhou. A taxi driving fraud detection system. In *2011 IEEE 11th International Conference on Data Mining*, pages 181–190. IEEE, 2011.

[17] E. Haertel. An application of latent class models to assessment data. *Applied Psychological Measurement*, 8(3):333–346, 1984.

[18] W. K. Hastings. Monte carlo sampling methods using markov chains and their applications. *Biometrika*, 57(1):97–109, 1970.

[19] G.-J. Hwang and S.-Y. Wang. Single loop or double loop learning: English vocabulary learning performance and behavior of students in situated computer games with different guiding strategies. *Computers & Education*, 102:188–201, 2016.

[20] J. Johns and B. Woolf. A dynamic mixture model to detect student motivation and proficiency. In *Proceedings of the National Conference on Artificial Intelligence*, 2006.

[21] D. H. Jonassen. *Handbook of research on educational communications and technology*. Taylor & Francis, 2004.

[22] B. W. Junker and K. Sijtsma. Cognitive assessment models with few assumptions, and connections with nonparametric item response theory. *Applied Psychological Measurement*, 25(3):258–272, 2001.

[23] J. Kivinen and M. K. Warmuth. Additive versus exponentiated gradient updates for linear prediction. In *Proceedings of the twenty-seventh annual ACM symposium on Theory of computing*, pages 209–218. ACM, 1995.

[24] A. Klementiev, D. Roth, and K. Small. An unsupervised learning algorithm for rank aggregation. In *European Conference on Machine Learning*, pages 616–623. Springer, 2007.

[25] Y. Koren. Factorization meets the neighborhood: a multifaceted collaborative filtering model. In *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 426–434. ACM, 2008.

[26] D. D. Lee and H. S. Seung. Algorithms for non-negative matrix factorization. In *Advances in NIPS*, pages 556–562, 2001.

[27] G. Linden, B. Smith, and J. York. Amazon. com recommendations: Item-to-item collaborative filtering. *Internet Computing, IEEE*, 7(1):76–80, 2003.

[28] A. Mnih and R. Salakhutdinov. Probabilistic matrix factorization. In *Advances in NIPS*, pages 1257–1264, 2007.

[29] G. Rasch. On general laws and the meaning of measurement in psychology. In *Proceedings of the fourth Berkeley symposium on mathematical statistics and probability*, volume 4, pages 321–333. University of California Press Berkeley, CA, 1961.

[30] P. Resnick, N. Iacovou, M. Suchak, P. Bergstrom, and J. Riedl. Grouplens: an open architecture for collaborative filtering of netnews. In *Proceedings of the 1994 ACM conference on Computer supported cooperative work*, pages 175–186. ACM, 1994.

[31] T. A. Schonhoff and A. A. Giordano. *Detection and estimation theory and its applications*. Pearson College Division, 2006.

[32] J. A. Swets. Signal detection and recognition in human observers: Contemporary readings. 1964.

[33] J. B. Sympson. A model for testing with multidimensional items. In *Proceedings of the 1977 computerized adaptive testing conference*, pages 82–98. University of Minnesota, Department of Psychology, Psychometric Methods Program Minneapolis, 1978.

[34] B. E. Vaessen, F. J. Prins, and J. Jeuring. University students' achievement goals and help-seeking strategies in an intelligent tutoring system. *Computers & Education*, 72:196–208, 2014.

[35] W. J. van der Linden and R. K. Hambleton. *Handbook of modern item response theory*. Springer Science & Business Media, 2013.

[36] M. Volkovs and R. S. Zemel. New learning methods for supervised and unsupervised preference aggregation. *Journal of Machine Learning Research*, 15(1):1135–1176, 2014.

[37] J. A. Walonoski and N. T. Heffernan. Detection and analysis of off-task gaming behavior in intelligent tutoring systems. In *International Conference on Intelligent Tutoring Systems*, pages 382–391. Springer, 2006.

[38] L. Wu, E. Chen, Q. Liu, L. Xu, T. Bao, and L. Zhang. Leveraging tagging for neighborhood-aware probabilistic matrix factorization. In *Proceedings of the 21st ACM international conference on Information and knowledge management*, pages 1854–1858. ACM, 2012.

[39] R. Wu, Q. Liu, Y. Liu, E. Chen, Y. Su, Z. Chen, and G. Hu. Cognitive modelling for predicting examinee performance. In *Proceedings of the 24th International Joint Conference on Artificial Intelligence*, pages 1017–1024. AAAI Press, 2015.

[40] H. Zhu, H. Xiong, Y. Ge, and E. Chen. Discovery of ranking fraud for mobile apps. *IEEE Transactions on Knowledge and Data Engineering*, 27(1):74–87, 2015.