Conceptual Issues in Mastery Criteria: Differentiating Uncertainty and Degrees of Knowledge

Radek Pelánek

Masaryk University Brno
Czech Republic
pelanek@fi.muni.cz

Abstract. Mastery learning is a common personalization strategy in adaptive educational systems. A mastery criterion decides whether a learner should continue practice of a current topic or move to a more advanced topic. This decision is typically done based on comparison with a mastery threshold. We argue that the commonly used mastery criteria combine two different aspects of knowledge estimate in the comparison to this threshold: the degree of achieved knowledge and the uncertainty of the estimate. We propose a novel learner model that provides conceptually clear treatment of these two aspects. The model is a generalization of the commonly used Bayesian knowledge tracing and logistic models and thus also provides insight into the relationship of these two types of learner models. We compare the proposed mastery criterion to commonly used criteria and discuss consequences for practical development of educational systems.

Keywords: learner modeling · mastery learning · uncertainty

1 Introduction

A common approach to personalization in educational systems is mastery learning, which is an instructional strategy that requires learners to master a topic before moving to more advanced topics [15]. A key aspect of mastery learning is a mastery criterion – a rule that determines whether a learner has achieved mastery [13]. A typical application of mastery criterion within a modern educational system is the following. A learner solves problems or answers questions in the system. Data about learner performance are summarized by a model of learner knowledge or by some summary statistic. This number is compared to a threshold that specifies the strictness of the criterion. A simple example of such a criterion is "N consecutive correct" – performance is summarized as the number of correctly answered question in a row and the threshold is a natural number N. The choice of the threshold involves a trade-off between unnecessary over-practice and the risk of premature declaration of mastery.

But what exactly is the meaning of the threshold? Does it specify how large portion of the practiced topic the learner has mastered? Or does it specify how certain can we be that the learner has sufficiently mastered the topic? The main point of this work is that it is useful to explicitly differentiate these two aspects of mastery: degrees of knowledge and certainty of knowledge estimation.

The current state-of-the-art makes the issues of threshold choice and interpretation obfuscated. There are two influential families of learner models: models based on Bayesian knowledge tracing (BKT) and family of logistic models [12]. These approaches estimate knowledge using different assumptions, which leads to significantly different interpretations of mastery thresholds.

The BKT model [3] makes a key simplifying assumption that knowledge can be modeled as a binary state (known/unknown). The model provides a probability estimate quantifying the certainty that the learner is in the known state. This estimate is commonly compared to a threshold 0.95, which leads to a clear interpretation that there is 95% chance that the learner has already mastered the topic. This interpretation, however, holds only under the idealistic assumption of binary knowledge. The assumption may be reasonable for very fine-grained knowledge components with homogeneous items (e.g., "addition of fractions with the same denominator"). In many practical cases, however, learner models are applied to more coarse-grained knowledge components, where the assumption of binary knowledge is far from satisfied. This degrades the performance of the model, and – for our discussion more importantly – it obfuscates the interpretation of the model estimate and the threshold. In these cases the BKT incorporates the degrees of knowledge aspect into the skill estimate, which thus loses the clear probabilistic interpretation.

Another common learner modeling approach is the family of models based on the logistic function, e.g., Rasch model, Performance factor analysis [10], or the Elo rating system [11]. These models utilize assumption of a continuous latent skill θ and for the relation between the skill and the probability of correct answer use the logistic function $\sigma(\theta) = \frac{1}{1+e^{-\theta}}$. The skill is estimated based on the learner's performance, for example in the Performance factor analysis model the skill is given by a weighted combination of correct and incorrect responses. These models thus utilize continuous knowledge scale, but do not explicitly quantify the uncertainty of estimates – they typically provide only a point estimate of the skill, which combines uncertainty and knowledge estimate into one number.

To address these issues, we propose a relatively simple model that generalizes both BKT and logistic model and provides clear differentiation of the degrees of knowledge and the uncertainty of estimates. The model is a special case of the hidden Markov model that uses a logistic function for specifying emission probabilities. The model leads to a conceptually clear mastery criterion with two thresholds. The first threshold specifies what degree of knowledge we consider to be sufficient for mastery. The second threshold specifies the degree of uncertainty we are willing to tolerate in the mastery decision. We then compare this criterion to other mastery criteria and discuss relation to previous work.

2 Modeling Uncertainty and Degrees of Knowledge

The BKT model provides a conceptual treatment of uncertainty, but does not address the degrees of knowledge, whereas logistic models model different degrees of knowledge, but do not address uncertainty in systematic way. We can address both these aspects by using a more general hidden Markov model than BKT.

2.1 LogisticHMM Model

A general Hidden markov model (HMM) models a process with a discrete latent state and noisy observations. A state at time t is denoted q_t, observation is denoted O_t. An HMM has the following elements:

- discrete set of latent states $\{s_0, s_1, \dots, s_{n-1}\}$ of size n,
- discrete set of of observation $\{o_0, o_1, \dots, o_{m-1}\}$ of size m,
- transition probabilities $T_{ij} = P(q_{t+1} = s_j | q_t = s_i)$ (a matrix $n \times n$ with rows summing to 1),
- emission probabilities $E_{ij} = P(O_t = o_j | q_t = s_i)$ (a matrix $n \times m$ with rows summing to 1),
- initial state probabilities: $\pi_i = P(q_1 = s_i)$.

For discussion of mastery criteria we propose "LogisticHMM" – a special version of the HMM that is general enough to generalize both the BKT and logistic models, and yet specific enough to be practically applicable (e.g., it has a small number of parameters). When modeling knowledge, the n latent states correspond to skill modeled with n degrees of knowledge. The observations correspond to learners answers to questions. We consider only the basic case with two observations: correct and incorrect.

Transition probabilities model learning. The basic version is to consider a single parameter l (speed of learning) and define the transition function as:

- $T_{ii} = 1 - l$ (learning did not occur and learner states in the same state),
- $T_{i(i+1)} = l$ (a learner learned and moves to the next knowledge states),
- T_{ij} is zero in all other cases (i.e., we model neither forgetting, nor sudden large increases in knowledge).

The emission probabilities are specified using the logistic function:

$$P(correct|s_i) = \sigma(a(i/(n-1) - b)) \quad (1)$$

where $\sigma(x) = 1/(1 + e^{-x})$ is the logistic function and a, b are parameters that specify the difficulty b and discrimination a of the modeled knowledge component. Fig. 1 provides an illustration of the emission probabilities.

The model can be easily generalized. For example, we can model forgetting or more important increases in knowledge in transition probabilities. For observations we can consider more general distribution functions over more fine-grained observations, e.g., using partial credit scoring, utilizing response times, or taking into account difficulty of individual items. Distribution of the initial skill estimate can be used to incorporate information from other KCs (prerequisite skills). However, for our discussion of relations of modeling and mastery learning the presented version is sufficient.

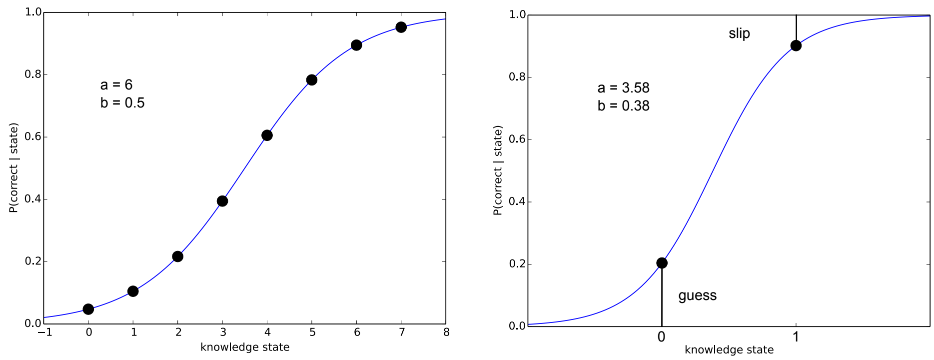

Fig. 1. Examples of emission probabilities in the LogisticHMM model. Left: A running example used throughout the paper. Right: An example illustrating that BKT is a special case of the LogisticHMM model.

2.2 BKT and Logistic Models as Special Versions

BKT and logistic models can be seen as special versions of the model or closely approximated by special versions of the model. The standard Bayesian knowledge tracing [3] is a hidden Markov model where skill is a binary latent variable (either learned or unlearned). The model has 4 parameters: P_i is the probability that the skill is initially learned, P_l is the probability of learning a skill in one step, P_s is the probability of an incorrect answer when the skill is learned (slip), and P_g is the probability of a correct answer when the skill is unlearned (guess). This can be expressed a special case of our LogisticHMM for $n = 2$. The probability that the skill is initially learned and the probability of learning a skill in one step remain the same. The slip and guess parameters can be transformed into a, b parameters without loss of generality. Fig. 1 (right) provides illustration of the basic relation. By substituing $n = 2$ and $i = 0, i = 1$ into equation 1 we get: $P_g = \sigma(-ab); (1 - P_s) = \sigma(a(1 - b))$. From this we can solve for a, b:

$$a = logit(1 - P_s) - logit(P_g)$$
$$b = -logit(P_g)/a$$

where *logit* is the inverse of logistic function, i.e., $logit(p) = \log((1 - p)/p)$.

Logistic models utilize continuous skill, but from practical perspective the difference between continuous skill and discrete skill with large n is negligible. Even $n = 10$ should be in most practical cases sufficient, since the model clearly involves other, more important simplification with respect to the reality, and thus the difference between continuous skill and its discretized approximation is not a fundamental one.

2.3 Using the Model in Mastery Criterion

Given a sequence of observations, the skill estimates can be computed using the standard forward algorithm for HMMs, i.e., using the Bayes theorem for computing the posterior distribution after the observation and updating it with the transition probability. Specifically, for our case the computation can be performed as follows. The skill estimate after observing t answers is $\boldsymbol{p}_t$ – a vector of length n with a sum one; p_{ti} is the probability of student being in the state i at time t. Initial estimate is given by the initial state probabilities: $\boldsymbol{p}_0 = \pi$. The estimate for $\boldsymbol{p}_{t+1}$ is computed from the estimate $\boldsymbol{p}_t$ by:

1. Taking into account the observation o_t at $t+1$ we compute an auxiliary estimate $\boldsymbol{p}'_{t+1}$:
$$p'_{(t+1)i} \propto p_{ti} L(o_t|s_i)$$
where $L(o_t|s_i)$ is the likelihood of observing a given answer, i.e., $\sigma(a(i/(n-1)-b))$ for a correct answer and $1-\sigma(a(i/(n-1)-b))$ for an incorrect answer.
2. Updating $\boldsymbol{p}'_{t+1}$ by transition probabilities we obtain the new estimate $\boldsymbol{p}_{t+1}$:
$$p_{(t+1)i} \propto \sum_j p'_{(t+1)j} T_{ji}$$

Both vectors $\boldsymbol{p}'_{t+1}$ and $\boldsymbol{p}_{t+1}$ need to be normalized to sum to 1. Note that the procedure is just a slightly more general version of the commonly used procedure for computing skill estimate under the Bayesian knowledge tracing model [12].

Fig. 2 shows a specific illustration using our running example from Fig. 1 with the learning speed $l = 0.3$. The initial distribution is depicted in the first graph; the other three graphs show the estimated skill distribution after 5th, 10th, and 19th answer.

With this model we can do mastery detection with systematic treatment of uncertainty and degrees of knowledge. The mastery criterion has two parameters:

- threshold T_m specifying which state is sufficient for mastery,
- threshold T_u specifying how certain we want to be.

We declare mastery when the probability that the skill is larger than T_m is larger than T_u, i.e., $P(\theta \geq T_m) \geq T_u$. For the example in Fig. 2 with $T_m = 6$ and $T_u = 0.95$ mastery is declared after the 19th answer.

3 Comparison with Other Mastery Criteria

The presented model provides a conceptually clear approach to thresholds in mastery criteria. Is this just a conceptual tool for thinking about mastery, or does the usage of the model also lead to practically important differences in mastery decisions?

Evaluation of mastery criteria is very complex. For data from real systems we do not have "ground truth", so it is difficult to perform fair comparison of different criteria [13]. Moreover, the performance of criteria interacts with issues like parameter fitting of used models. Therefore, to explore our question we perform experiments with simulated data – comparing the behavior of criteria

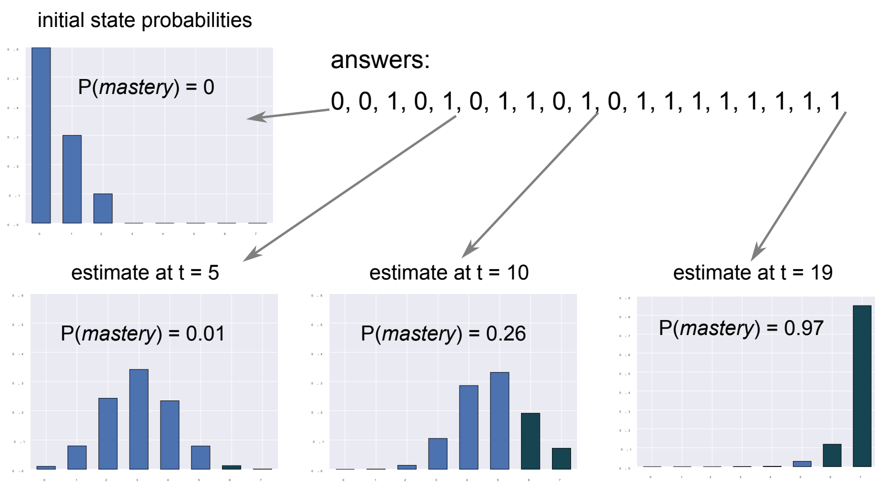

Fig. 2. Example of model estimates for a sample sequence of answers. The diagrams show estimated skill before the first attempt (i.e., the initial distribution π) and after the 5th, 10th, and 19th attempts. The probability of mastery is computed for the mastery state 6 ($T_m = 6$).

based on the LogisticHMM model with other commonly used criteria. We use the following approach. Using the LogisticHMM model we generate simulated learner data. For the generated data we perform mastery decision using both the LogisticHMM model and other mastery criteria. We compare the decision to the ground truth and analyze their agreement.

This is, of course, a simplified setting. It is also optimistic for the LogisticHMM model, since we avoid the issue of parameter fitting and use the same model parameters for generating the data and making the mastery decisions. Our purpose, however, is not to evaluate whether the LogisticHMM *is* better than other criteria, but whether it *could* be better and under what circumstances.

3.1 Comparison with N Consecutive Correct Criterion

If the noise in observations is low (parameter a in the LogisticHMM model is high), then even the the basic N consecutive correct (NCC) criterion leads to good estimates. By a suitable choice of N in the NCC criterion we can achieve nearly the same decisions as by the LogisticHMM model. In these cases the advantage of the LogisticHMM model is only the better interpretability of threshold parameters.

The NCC criterion has worse performance when observations are noisy (parameter a in the LogisticHMM model is low). Intuitively, this is because the NCC criterion has very limited "memory" – once incorrect answer is observed, the counting of correct answers starts from zero and all information about previous attempts is forbidden. When observations are noisy, incorrect answers are

Table 1. Best EMA parameters for different setting of threshold parameters for the running example from Figure 1 (with $l = 0.1$ and the initial state distribution same as in Fig. 2).

Best fitting parameter α

$T_m \setminus T_u$	0.6	0.7	0.8	0.9	0.95	0.98
5	0.85	0.85	0.9	0.85	0.9	0.9
6	0.85	0.85	0.9	0.9	0.85	0.9
7	0.9	0.85	0.9	0.9	0.95	0.9

Best fitting parameter T

$T_m \setminus T_u$	0.6	0.7	0.8	0.9	0.95	0.98
5	0.65	0.7	0.65	0.8	0.75	0.8
6	0.8	0.85	0.8	0.85	0.95	0.92
7	0.85	0.92	0.9	0.92	0.85	0.97

relatively common even for skilled learners and thus it is advantageous to remember a summary of previous performance, which the LogisticHMM model does using the estimated skill distribution.

3.2 Comparison with Exponential Moving Average

Simple, yet flexible and pragmatically advantageous approach to mastery decision is the exponential moving average (EMA) approach [13]. With this approach we compute the exponentially weighted average of past answers and compare this statistics to a threshold. The moving average can be, of course, consider also with other weights; the exponential weighting has the advantage that it can be computed without the need to store and access the whole history of learners attempts. It can be computed using the update rule $\theta_k = \alpha \cdot \theta_{k-1} + (1 - \alpha) \cdot c_k$, where θ_k is the skill estimate after k-th answer, c_k is the correctness of the k-th answer, and α is a discounting parameter, which controls the relative weight of recent attempts. The mastery criterion is the basic comparison to a threshold: $\theta_k \geq T$.

This approach overcomes the NCC limitation of "limited memory" – the skill estimate θ_k now summarizes the whole history of answers, just giving the more recent attempts more weight. The approach has two parameters: α and T. Results of experiments with simulated data suggest that by tuning these parameters the EMA approach can quite well imitate the LogisticHMM – the correlation of their decisions is often over 0.9. However, the exact setting and interpretation of these parameters is not straightforward. For example, the optimal value of α depends not just on both the requested degree of mastery and level of uncertainty, but also on the the expected speed of learning (if learning is fast, α should be low, so that the estimate is not influenced much by old data).

Table 1 provides a specific illustration. For the LogisticHMM illustrated in Figure 1 (with $l = 0.1$), we have fitted the optimal values of α and T of the EMA approach. The values were fitted using a grid search while trying to optimize the agreement of mastery decision between the LogisticHMM model and the EMA approach (for 1000 simulated learners). As we can see, both these parameters change depending on both T_m and T_u, i.e., they are not easily linked to the interpretable parameters of the LogisticHMM approach.

3.3 Comparison with Bayesian Knowledge Tracing

The BKT model is typically used for detecting mastery using the basic threshold policy. The model estimates the probability of being in the learned states and declares mastery when this probability is over a given threshold T (with the commonly used value 0.95). Under the assumption of binary knowledge state, this method has a clear interpretation – it uses a threshold on the uncertainty of the skill estimate. However, when the learner performance is based on multiple degrees of knowledge, the BKT model estimate starts to combine both uncertainty and degrees of knowledge and becomes hard to interpret.

As a specific example, consider our running example from Figure 1 (with $l = 0.3$) and let us use the BKT model to do mastery decisions for this case. At first, we need to estimate the BKT parameters. Note that the fitted parameters of BKT depend on the number of attempts per learner that we use for fitting (this aspect has been previously noted [14], but mostly it is not taken into account in literature):

	init	learn	guess	slip
15 attempts per learner	0.01	0.10	0.12	0.31
50 attempts per learner	0.00	0.11	0.18	0.07

In both cases we can notice the relatively high guess and slip parameters. These do not correspond to the behaviour of learners – according to the model that generated the data, learners who really know the topic have very high probability of answering correctly. The high guess and slip parameters are due to the simplifying binary assumption of BKT. BKT parameters reported in research paper often have high guess and slip parameters. As this example illustrates, the reason for this may be inappropriateness of the model assumptions rather than high propensity of students to slip or guess.

If we use the common 0.95 threshold under these circumstances, BKT model leads to premature declaration of mastery for all learners. Fig. 3 provides illustration that shows the results for our running example. Each dot corresponds to one simulated learner. The x axis shows the ground truth mastery (when did the learner reach the mastery state). The y axis shows when the mastery was declared according to the BKT model and according to the LogisticHMM (which uses the same parameters as the ground truth model, but estimates mastery from the generated noisy observations). Both the LogisticHMM and BKT model use the threshold 0.95 on uncertainty. As we can see, the LogisticHMM mastery estimates lag behind the ground truth moments of mastery. This is necessary – if we want to detect mastery with reasonable certainty from noisy observations, there is a necessary lag. The BKT model, on the other hand, nearly always declares mastery too early, which is an unwelcomed behaviour that is caused by the unsatisfied assumptions of the model.

Fig. 3. Comparison of mastery decisions made by the LogisticHMM model and th BKT model with the threshold on uncertainty 0.95.

4 Discussion and Related Work

We presented a proposal for conceptual approach to mastery learning based on a specific version of the hidden Markov model and compared the mastery criterion based on this model to several common mastery criteria. We now discuss consequences of this work to practical applications and relations to previous research.

4.1 Consequences for Practice

Although we present a new learner model, the point is not that this model should be used in implementations of educational systems. The use of the model would require additional steps, for example the choice of the number of states and parameter estimation, which we have not discussed. Based on our analysis of simulated data and on the general experience in the field of educational data mining, we do not expect that the model would bring a significant improvement in predictive accuracy or significantly different mastery decisions in real applications.

The main point of the model is that it provides conceptually clear way to think about mastery criteria and thresholds. Our experiments with mastery criteria suggest that simple criteria (specifically exponential moving average) may be expressive enough for practical purposes. However, such pragmatic approaches have parameters, which do not have clear interpretation and which are hard to set. We propose the LogisticHMM model primarily as tool for clarification and better usage of such more pragmatic approaches.

One specific clarification that our analysis brings concerns the interpretation of BKT in mastery learning. Based on the current research literature, one may be tempted to use the BKT model, fit it to the available data and use it for mastery decisions with a threshold 0.95 with an interpretation "mastery is declared when there is 95% certainty that the learner mastered the skill". This interpretation is, however, misleading. Unless the data correspond very closely to the strong assumptions of the BKT model, the interpretation is incorrect (possibly quite seriously).

The pragmatic conclusion for developers of educational systems can be formulated as follows: Instead of using a complex model that presents summary of knowledge as a single number, it may be preferable to utilize a simple approach that explicitly separates the estimated degree of knowledge the uncertainty of the estimate. Even using statistics like "average performance" and "number of attempts" may be in some cases sufficient. Such approach may lack statistical sophistication, but it can possibly provide similar decisions as complex models, clearer interpretability, and easier setting of thresholds.

4.2 Learner Modeling

Another purpose of the proposed LogisticHMM model is that it connects and clarifies previous research on learner modeling using BKT and logistic models. Our discussion highlights the simplifying assumption of binary knowledge state in the BKT model. On one hand, this assumption is quite clear and explicit in the model and was appropriate for the original use of the model in cognitive tutors [3]. However, the model is now widely used, often in situations where the assumption of binary knowledge is not appropriate. Previous work has proposed BKT extension with multiple states ("spectral BKT") [4], but such work is currently marginal in the field.

Recently, several works have tried to combine BKT and models based on logistic function. This has been done specifically in two closely related approaches: one based on incorporation of features to knowledge tracing [5], another using integration of latent-factors and knowledge tracing [8]. Although these approaches enable inclusion of "learners ability" into BKT and utilize the logistic function for predictions, they still keep the basic BKT assumption of binary knowledge state. Moreover, these works focused on predicting learner performance and not on mastery criteria.

A fully conceptual approach to treatment of uncertainty and skill estimation is to use dynamic Bayesian networks [2, 6], which can be used to capture relations between skill. Bayesian networks provide clean conceptual approach, but are difficult to apply in a practical implementation (e.g., it is difficult to specify or fit model parameters). Existing realization thus typically revert to simplifying assumptions, specifically to the assumption of binary knowledge state for each skill.

Bayesian methods in connection with mastery criteria have been used in computerized adaptive testing [9, 18]. But the context of testing has several differences from the context of online educational systems, the main difference is learning – in the context of testing the models typically assume that knowledge is not changing during the test; in educational systems the main goal is to increase knowledge during practice and thus the change of knowledge is a fundamental aspect of models.

4.3 More General Stopping Criteria

In this work we have studied the problem of stopping practice in the case of mastery. The stopping problem can be formulated more generally – we may want to stop the practice once asking more questions does not have any merit. Specifically, in addition to mastery we may want to detect "wheel-spinning learners" who are unable to master a topic [1]. Previous research have proposed criteria that can deal with this more general problem: instructional policies called predictive similarity [16] and predictive stability [7]. These works, however, pay little attention to the choice of thresholds. Specifically, they add a parameter ϵ, which specifies the size of the difference in model predictions that is considered large enough to warrant further practice. It is not clear how to set and interpret this parameter.

A more conceptual approach to the general stopping problem could be mixture modeling, which has been previously proposed for modeling individual learning curves [17]. We can use a mixture model, with one component specifying learners who are improving and are able to eventually reach mastery and a second component specifying wheel-spinning learners. Using such model we can estimate the probability that a learner belongs to each class. This leads to a stopping criterion where thresholds are probabilities with clear interpretation.

References

1. Beck, J.E., Gong, Y.: Wheel-spinning: Students who fail to master a skill. In: Proc. of Artificial Intelligence in Education. pp. 431–440. Springer (2013)
2. Conati, C., Gertner, A., Vanlehn, K.: Using bayesian networks to manage uncertainty in student modeling. User modeling and user-adapted interaction **12**(4), 371–417 (2002)
3. Corbett, A.T., Anderson, J.R.: Knowledge tracing: Modeling the acquisition of procedural knowledge. User modeling and user-adapted interaction **4**(4), 253–278 (1994)
4. Falakmasir, M., Yudelson, M., Ritter, S., Koedinger, K.: Spectral bayesian knowledge tracing. Proc. of Educational Data Mining pp. 360–363 (2015)
5. González-Brenes, J., Huang, Y., Brusilovsky, P.: General features in knowledge tracing: Applications to multiple subskills, temporal item response theory, and expert knowledge. In: Proc. of Educational Data Mining. pp. 84–91 (2014)
6. Kaeser, T., Klingler, S., Schwing, A.G., Gross, M.: Dynamic bayesian networks for student modeling. IEEE Transactions on Learning Technologies (2017)
7. Käser, T., Klingler, S., Gross, M.: When to stop?: Towards universal instructional policies. In: Proc. of Learning Analytics & Knowledge. pp. 289–298. ACM (2016)
8. Khajah, M., Wing, R.M., Lindsey, R.V., Mozer, M.C.: Integrating latent-factor and knowledge-tracing models to predict individual differences in learning. In: Proc. of Educational Data Mining (2014)

9. Lewis, C., Sheehan, K.: Using bayesian decision theory to design a computerized mastery test. Applied Psychological Measurement **14**(4), 367–386 (1990)
10. Pavlik, P.I., Cen, H., Koedinger, K.R.: Performance factors analysis-a new alternative to knowledge tracing. In: Proc. of Artificial Intelligence in Education. pp. 531–538. IOS Press (2009)
11. Pelánek, R.: Applications of the elo rating system in adaptive educational systems. Computers & Education **98**, 169–179 (2016)
12. Pelánek, R.: Bayesian knowledge tracing, logistic models, and beyond: an overview of learner modeling techniques. User Modeling and User-Adapted Interaction **27**(3), 313–350 (2017)
13. Pelánek, R., Řihák, J.: Experimental analysis of mastery learning criteria. In: Proc. of User Modelling, Adaptation and Personalization. pp. 156–163. ACM (2017)
14. Pelánek, R., Řihák, J., Papoušek, J.: Impact of data collection on interpretation and evaluation of student model. In: Proc. of Learning Analytics & Knowledge. pp. 40–47. ACM (2016)
15. Ritter, S., Yudelson, M., Fancsali, S.E., Berman, S.R.: How mastery learning works at scale. In: Proc. of ACM Conference on Learning@Scale. pp. 71–79. ACM (2016)
16. Rollinson, J., Brunskill, E.: From predictive models to instructional policies. In: Proc. of Educational Data Mining. pp. 179–186 (2015)
17. Streeter, M.: Mixture modeling of individual learning curves. In: Proc. of Educational Data Mining. pp. 45–52 (2015)
18. Vos, H.J.: A bayesian procedure in the context of sequential mastery testing. Psicológica **21**(1), 191–211 (2000)