

Bayesian Knowledge Tracing, Logistic Models, and Beyond: An Overview of Learner Modeling Techniques

Radek Pelánek

Accepted; preprint version

Abstract Learner modeling is a basis of personalized, adaptive learning. The research literature provides a wide range of modeling approaches, but it does not provide guidance for choosing a model suitable for a particular situation. We provide a systematic and up-to-date overview of current approaches to tracing learners' knowledge and skill across interaction with multiple items, focusing in particular on the widely used Bayesian knowledge tracing and logistic models. We discuss factors that influence the choice of a model and highlight the importance of the learner modeling context: models are used for different purposes and deal with different types of learning processes. We also consider methodological issues in the evaluation of learner models and their relation to the modeling context. Overall, the overview provides basic guidelines for both researchers and practitioners and identifies areas that require further clarification in future research.

Keywords Learner modeling · Skill modeling · Overview · Evaluation · Methodology · Knowledge-learning-instruction framework

1 Introduction

Learner modeling is a key element in adaptive educational systems. Based on the available observational data about a learner's interaction with an educational system, a learner model provides an estimate of the current state of a learner. This estimated state is used to guide the personalized behavior of an adaptive system.

Learner modeling has received significant attention from the scientific community and today encompasses a wide range of constructs including cognitive skills, affect, motivation, and meta-cognition. In this article we focus on modeling knowledge and cognitive skills – the most typical type of learner modeling

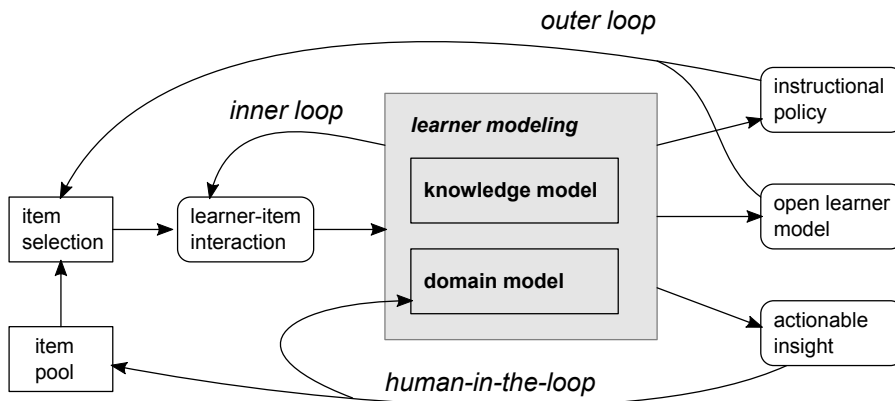


Fig. 1 The context of learner modeling.

as well as being the fundamental model type for adaptive learning. Modeling of knowledge and skills is concerned with what the learner knows and is able to do. Such models estimate the current state of a learner’s knowledge and provide a prediction of future performance. Knowledge modeling is inextricably linked to modeling the content of a particular domain, e.g., the definition of knowledge components, the mapping of specific items to these knowledge components, and the modeling of item difficulty.

Fig. 1 illustrates a typical context of modeling learners’ knowledge: a learner interacts with an item; data about her performance are processed by the model; the model outputs (e.g., knowledge estimates, predictions of future performance) are used to influence her future interactions with the system (within one of the “loops”). As the figure illustrates, learner models can be used for several different purposes. The “inner loop” is concerned with a learner’s short-term behavior in the process of interacting with a single item. Other loops are concerned with tracing a learner’s knowledge across multiple items. In this work we focus mainly on models with this long-term perspective, which include the extensively studied families of Bayesian knowledge tracing models and logistic models.

Even this specific area of learner modeling encompasses a wide range of techniques and issues. Recent intensive research has made considerable progress that has sparked a lot of innovative ideas in learner modeling. Thanks to this progress, there are now many aspects that we may elaborate in a learner model:

- modeling the processes of learning and forgetting (e.g., Bayesian knowledge tracing, logistic models, neural networks),
- the use of different types of observational data (e.g., correctness of answers, response times, the use of hints, wrong answers),
- domain modeling (e.g., a definition of knowledge components and their relations, a mapping of items to knowledge components),
- learner clustering, the level of individualization.

Learner modeling involves many decisions – we need to decide which aspects to include in our model and how exactly to model them. The current research, unfortunately, provides only little guidance for making these choices. This is a problem for both researchers and developers. For researchers, it is difficult to evaluate the merit of newly proposed techniques. Since there are many potential candidates, it is difficult for a researcher to choose the most suitable representatives for comparison. For developers of real-world applications, there is little practical guidance for choosing and tuning a learner model for a particular application. This ultimately hinders the spread of intelligent techniques.

There cannot be any simple, universal advice for making the choices, because learner modeling has a wide range of applications. A different modeling approach is necessary for a model used for adaptively ordering foreign language vocabulary and for a model used for the discovery of a suitable structure of high-school algebra. The proper choice of a learner model depends on the specific context of learner modeling:

- the type of knowledge being modeled (e.g., simple facts, categorization rules, or complex skills),
- the input data available for modeling (what kinds of activities are available in a given educational system and what data about learners can be collected),
- the purpose of the model, i.e., how will the model results be used (e.g., as a part of an automated instructional policy, visualized as an open learner model, or rather serve as a basis for manual intervention based on actionable insight).

To make sense of the wide learner modeling landscape, it is necessary to take such contextual information into account.

An important part of learner modeling is model fitting, selection, and evaluation. There are also many possibilities within each of these steps. Typically, we can choose among several parameter fitting procedures, metrics for measuring model quality, cross-validation approaches, and other methodical steps. As is the case with the choice of the model itself, there are no simple answers or universal evaluation approaches. The choice of evaluation methodology again depends on the context of modeling.

This overview provides an up-to-date and systematic overview of major trends in modeling learners' knowledge and skill. Although overviews of learner modeling are available in previous works (Desmarais and Baker, 2012; Chrysafiadi and Virvou, 2013; Pavlik et al., 2013), this overview contributes in several directions:

- Previous overviews consider learner modeling in general while this overview considers only the modeling of knowledge and skills. With this constraint, we can provide a more detailed and nuanced discussion of relevant techniques.
- We provide a discussion of recent work. Modeling of knowledge and skills has received significant attention in recent research – more than 60 papers

included in this overview have been published since 2014 and are thus not covered in previous overviews.

- We discuss the purposes of models, e.g., an automatic guidance of system behavior, open learner modeling, and actionable insight. We describe how the purpose of a model impacts choices made in learner modeling. Dedicated discussions of such contexts are as rare in individual research studies as they are in previous overviews.
- We study connections between learner modeling and types of learning processes relevant to a particular application (e.g., memory, induction, understanding). These connections can help us understand the relative merits of different modeling techniques.
- We pay significant attention to methodological issues concerning model fitting and evaluation. We discuss the impact of the purpose of the model on decisions made in the processes of parameter fitting, comparison, and the interpretation of models. These aspects are not covered in previous surveys (or only briefly).

To study the connection between learner modeling and relevant learning processes, we utilize the Knowledge-Learning-Instruction (KLI) framework (Koedinger et al., 2012). This framework was chosen because it addresses learning on a suitable level of granularity. The central focus of both the KLI framework and learner modeling are “knowledge components” (units of cognitive function). Frameworks that address learning on finer grained levels (e.g., theories of learning on a neural level) or coarser grained level (e.g., situated learning) cannot be easily connected to learner modeling techniques, at least not in the current state of research.

This overview is targeted at both developers and researchers. For developers, it should provide guidance in choosing and tuning a suitable learner model for a particular application. For researchers, it should provide help with placing their work in the landscape of existing research. Since the current state of learner modeling research does not fully address all questions about the suitability of modeling approaches in different contexts, we also propose some explicit hypotheses for verification in future research.

2 Terminology

This overview covers a wide scope of research, which often uses different terminology for the same or very similar notions. Therefore, we start by clarifying the terminology as used within the scope of this paper. We also introduce the relevant parts of the knowledge-learning-instruction framework to which we relate learner modeling in the following sections.

2.1 Types of Learner Models

The current research uses both “learner modeling” and “student modeling” with basically the same meaning. As the term “student” tends to imply a

formal educational setting, we prefer to use the term “learner”, which is more general and more appropriate in the case of open learning environments in particular. Nevertheless, this choice is basically only stylistic since it does not have any impact on the relevance of individual modeling techniques.

A more important terminological issue is a clarification of what falls under the term “learner modeling”. What aspects of a learner’s state do we model? The basic and most common type of models focuses on modeling the knowledge and skills of learners. In this work we focus only on models of this type – this is the primary type of models used in most adaptive educational systems and even for this class of models we have a wide range of issues to explore. Many other aspects of learners can be (and have been) modeled, e.g., affect, motivation, meta-cognition, learner preferences, or behaviors like gaming the system. Several previous overview papers include a wide discussion of different types of learner model, see for example Desmarais and Baker (2012); Chrysafidi and Virvou (2013); Nakic et al. (2015); Valdés Aguirre et al. (2016); Pavlik et al. (2013). Our focus is on a narrower and deeper discussion of skills and knowledge modeling.

The basic goal of skills and knowledge modeling is to estimate the learners current knowledge state and to predict future performance based on data about past performance. The specific type of model application determines what data are available and what should be the basic form of a model. There are two basic types of adaptive behavior in educational systems, most often called inner loop and outer loop (Vanlehn, 2006), while other authors use “problem solving and solution analysis tutors” and “curriculum sequencing” (Desmarais and Baker, 2012) or microadaptivity and macroadaptivity (Essa, 2016). In the inner loop the focus is on the learner’s activity while solving a single multi-step problem. In the outer loop the focus is on a sequence of independent items, using estimates of knowledge from past items to choose suitable future items. Originally, the focus of the community has been on models used in the inner loop, e.g., cognitive tutors (Anderson et al., 1995) and constraint based models (Mitrovic et al., 2001). These types of models, however, tend to be rather specific for a particular domain and difficult to develop (Aleven et al., 2006). Essa (2016) argues that microadaptivity (inner loop) should be regarded as the primary realm of an instructor while adaptive systems should focus on macroadaptivity (outer loop). A lot of recent focus has been on models primarily targeting the outer loop – these include the extensively studied families of Bayesian knowledge tracing models and logistic models. This is the main type of models that we discuss in this paper.

2.2 Learner and Domain Modeling

The modeling of learner knowledge and skills is inherently interconnected with the modeling of the educational domain concerned. Previous works have used the terms “learner modeling”, “skill modeling”, and “domain modeling” with overlapping meanings. In this work we use the following terminology:

Table 1 Basic categories of knowledge components, according to the KLI framework (Koedinger et al., 2012), simplified.

Knowledge components			Learning processes
Application condition	Response	Labels	
constant	constant	fact, association	fluency building
variable	constant	category, concept	induction, refinement
variable	variable	rule, principle	understanding, sense-making

- *domain modeling* – modeling concerned with the structure of the domain (e.g., items, knowledge components, and their relations),
- *knowledge modeling* – the modeling of the current knowledge of learners with respect to a particular domain model, including learning and forgetting processes,
- *learner modeling* – a general term encompassing the previous two.

Another terminological issue concerns the things that are presented to learners and their groupings. We use the generic term *item*, which can represent both simple questions and complex interactive problems. Although in some cases it is important to distinguish between a multiple-choice question and a multi-step problem, for our discussion a generic item view is sufficient. Items can also be instructional material such as text and video, but we will focus mainly on interactive items that assess learner knowledge. Items, which are tangible and specific, are typically linked to more abstract educational constructs. For these we use the term *knowledge components* (KC) – other common terms are skills, concepts, principles, schemata.

2.3 Knowledge Components and the KLI Framework

One of our goals is to connect learner modeling to the Knowledge-learning-instruction (KLI) framework, which is an instructional framework on the level of knowledge components (Koedinger et al., 2012). In this framework a knowledge component is defined as “an acquired unit of cognitive function or structure that can be inferred from performance on a set of related tasks”. The KLI framework is based on two taxonomies: a taxonomy of knowledge components and a taxonomy of learning processes. Table 1 presents an overview of these taxonomies and their relations. Note that our presentation of the KLI framework is simplified. We cover only those aspects that are directly relevant to our discussion of learner modeling and we highlight only the main ideas behind the framework; for a more detailed discussion we refer the reader to Koedinger et al. (2012).

The taxonomy of KCs is based on the condition of application and the learner’s response to assessment events concerning a KC. Both application conditions and responses can be either constant or variable.

The simplest KCs are applied under constant conditions and require constant response. Such KCs are typically called facts or associations; typical examples are second language vocabulary or geography facts. More complex KCs, called categories or concepts, require constant responses under variable conditions, e.g., English determiners and the applications of Pythagoras' theorem. The most complex KCs require variable responses under variable conditions, e.g., solving an algebra word problem or a programming exercise. Such KCs are often called principles, procedures, plans, or rules.

Learning processes are also classified into three main types (Koedinger et al., 2012):

- *Memory and fluency-building processes.* Non-verbal processes that strengthen memory and compile knowledge. They make knowledge more automatic and composed.
- *Induction and refinement processes.* Non-verbal processes like generalization, discrimination, classification, categorization, and schema induction. They make knowledge more accurate and suitably general.
- *Understanding and sense-making processes.* Explicit, verbally mediated learning in which learners attempt to understand or reason – this includes the comprehension of descriptions, self-explanation, and scientific discovery processes.

The basic relation between the KC taxonomy and the learning processes taxonomy is that more complex types of KCs require more complex learning processes. But the relation between learning processes and types of KCs is asymmetrical (as depicted in Table 1). Fluency building and induction are also relevant for rules and principles – in many cases it is not sufficient to understand a rule, but also to be fluent in its application (a typical example is single digit multiplication). On the other hand, understanding and sense making processes are not useful for learning facts.

Knowledge components can be considered on different levels of granularity. Table 2 provides examples of knowledge components from different domains, of different types, and different levels of granularity. Note that the granularity can influence the type of knowledge components – on a very fine level of granularity nearly everything becomes a “constant-constant” knowledge component (but not usually a useful one). Finer grained KCs lead to potentially more precise models, but also more complex ones with more parameters, fewer data per KC, and thus worse consistency of parameter estimates. The chosen granularity of KCs thus has a significant impact on learner modeling.

3 Context and Purpose of Learner Modeling

Fig. 1 provides a simplified view of the context of learner modeling in an educational system. Learner modeling provides several outputs that can serve different purposes – the figure illustrates some common uses of learner modeling. As illustrated, learner modeling is typically part of a “loop”, i.e., of an

Table 2 Examples: KC types and levels of granularity.

granularity	the type of knowledge component		
	constant-constant fact	variable-constant category	variable-variable rule
coarse	names and locations of countries	capitalization rules	fractions
medium	names and locations of African countries	capitalization of nouns	addition of fractions
fine	location of Egypt	capitalization of days of the week	addition of fractions with the same denominator

iterative process where learner modeling is one of the steps. The notions of “inner loop” and “outer loop” are commonly used in the literature on intelligent tutoring systems (Vanlehn, 2006). Based on recent developments, we have also included in the figure the notion of “human-in-the-loop” – cases where outputs of a learner model are interpreted by a human (a researcher, a system developer) and this interpretation has specific consequences for the educational system concerned (e.g., the manual update of a domain model or the addition of new items).

A learner model can be used in many ways. We outline four typical uses of models to illustrate the scope of potential applications. The purpose of the model determines which model outputs are important. Whereas instructional policies consider mainly model predictions, open learner modeling focuses on learner parameters, and discovery with models may focus on item parameters. Consequently, the purpose of the model has an impact on the treatment of identifiability issues. A model is non-identifiable when different parametrizations of the model provide the same predictions of learner performance. Identifiability may not matter much for the use of models in instructional policies, but may be very important for open learner modeling. This illustrates that the purpose of a model has direct consequences for the choice of modeling approach, parameter fitting procedures, and evaluation methodology.

3.1 Inner Loop

The inner loop can use the current knowledge estimate to personalize a learner’s activity while they are interacting with an item. The item can involve multiple solution steps and the interaction can involve feedback on progress, hints, explanations, or solution review. Based on the estimated knowledge, a system can provide personalized feedback, dynamically adjust the user interface, and provide adaptive scaffolding and personalized hints. Specific algorithmic decisions may concern if and when to actively propose (enable) hints and which hint to choose. Inner loop behaviors are discussed in more detail by Vanlehn

(2006); Aleven and Sewall (2016) provide an analysis of inner loop behaviors in a practical system.

The complexity of the inner loop depends on the type of relevant learning process:

- Memory and fluency building processes: Typically the item is just a simple question, the only feedback is information about correctness and, in the case of a mistake, a correct answer.
- Induction and refinement processes: Items are still rather simple, but in this case it is often reasonable to provide explanations for answers which can be personalized based on estimated knowledge.
- Understanding and sense-making processes: Knowledge can be still tested using simple items, but in this case it also makes sense to use more complex items, e.g., multi-step problems with feedback on the correctness of individual steps, available hints, etc. The interaction with learner modeling may be much richer than in the previous two cases.

The use of learner modeling to personalize the inner loop is thus relevant mainly for understanding and sense-making processes. Applications that utilize modeling in the inner loop were indeed developed chiefly for mathematics, programming, and physics, i.e., domains with significant focus on understanding and sense-making processes. Specific well-known systems deal with algebra (Anderson et al., 1995), SQL database query language (Mitrovic et al., 2001), and physics (Vanlehn et al., 2005).

Two common modeling methods for the inner loop are model tracing models and constraint-based models. Model tracing models (Corbett and Anderson, 1994; Anderson et al., 1995) are based on the ACT theory of skill acquisition. A model is specified as a set of production rules – low level cognitive steps encoded as “if-then” rules. Using these rules a model can monitor a learner’s progress through the problem, and the interaction with a user can thus be based on the current state of problem solving. Production rules may encode also “buggy rules”, common misconceptions of learners. Using these rules a tutoring system can provide appropriate feedback after mistakes.

Constraint-based models (Ohlsson, 1994; Mitrovic et al., 2001, 2007, 2003) specify a set of constraints that a problem solution must satisfy. Violating a constraint triggers an action such as a feedback message. In contrast to model tracing, constraint-based models typically do not attempt to explicitly model paths toward solutions. As a consequence, they are applicable to more open-ended problems.

Models used in the inner loop can be characterized as “short-term” models, whereas models used in the outer loop are “long-term” models (Mayo and Mitrovic, 2001). Short-term models are typically connected to long-term models, e.g., a short-term model specifies constraints and monitors their violations, whereas a long-term model estimates the probability that a learner understands individual constraints (Mayo and Mitrovic, 2001).

3.2 Instructional Policy

A typical application of learner modeling is concerned with choosing or proposing items that are then presented to a learner; this algorithm is often denoted as an instructional policy and is typically based on predictions provided by a learner model. A common approach to developing an instructional policy is based on the principle of mastery learning. This approach is today used on large scale, for example in Carnegie Learning systems (Ritter et al., 2016). Mastery learning policy focuses on a definition of a “stopping criterion” – learners solve items within a particular knowledge component, the system estimates their knowledge and presents new items until a stopping criterion is satisfied. Once mastery is declared, the learner progresses to the next knowledge component. A common stopping criterion is “95 % chance that a learner knows the next item”, but recently more complex criteria have been proposed (Rollinson and Brunskill, 2015; Käser et al., 2016). On the other hand, this instructional policy may be used even without a learner model, e.g., using a simple “ k correct in row” stopping criterion. The advantage that learner modeling brings to instructional policies is not yet completely clear and may depend on the type and the granularity of a specific knowledge component.

Another common instructional policy is adaptive item sequencing, where the algorithm selects and orders the items (within a particular KC), taking into account the predictions provided by a model. The policy typically takes into account not just the model prediction, but also other factors, e.g., a desired difficulty of questions or timing information. These factors can be combined using some heuristic scoring function (Papoušek et al., 2014; David et al., 2016). A more principled approach is to use a decision theoretic approach and try to maximize the utility of a selected item (Mayo and Mitrovic, 2001).

For some knowledge components (particularly facts), it is important to take into account forgetting. Instructional policies in these cases try to realize the spaced repetition learning principle (Pavlik and Anderson, 2005), typically in the form of scheduling policies based on estimated knowledge (Reddy et al., 2016; Pavlik and Anderson, 2008; Pavlik Jr et al., 2008).

3.3 Open Learner Model

An open learner model (Bull and Kay, 2007, 2010; Mitrovic and Martin, 2007) makes the representation of a fitted model available to learners. It provides a visualization of estimated knowledge, typically in the form of skillometers or progress bars, possibly also using more complex visualization techniques like hierarchical trees or conceptual graphs. The idea of opening a learner model to learners has several potential goals, e.g., promoting the meta-cognitive abilities of learners (reflection, self-monitoring), facilitating discussion between learners and teachers (or parents), supporting self-regulated learning by providing navigation to items most relevant for a particular learner, enhancing engagement by using open models for social comparison, or increasing learner trust

in an adaptive system. In the case of persuadable models (Bull et al., 2016), where learners can provide additional or corrective information, the opening of a model can also increase the accuracy of model data.

For open learning modeling the key aspect of a model is the estimate of a learner’s knowledge state. The visualization is typically structured in the same way as the domain model, it thus also relies heavily on the knowledge component structure of the learner model.

A consequence of opening a learner model to learners is the increased importance of the intuitive behavior of model outputs. For example, a learner would typically expect that after a correct answer a skill estimate will increase and after an incorrect answer a skill estimate will decrease. Although such behavior may seem straightforward, some models (particularly more complex models like neural networks or mixture models) are able to achieve good predictive accuracy without guaranteeing this behavior.

3.4 Actionable Insight and Discovery with Models

The previously discussed applications of learner models are “automatic”, i.e., model outputs directly influence user experience. But model outputs may also be useful for humans who use them to “manually” influence the system behavior or the learners. This line of application corresponds to Baker’s argument for “stupid tutoring system, intelligent humans” (Baker, 2016) – keeping systems relatively simple and focusing on the use of model outputs for guiding human decisions.

Learner models can provide interesting insight for tool developers, content authors, teachers, or educational researchers. Typical model aspects useful for such insight are item difficulties, item learning rates, knowledge component structure and relations, and learner clusters. To be useful, insights obtained from a model should not only be interesting, but also actionable, i.e., they should have specific consequences for further development of the educational system concerned or for relevant stakeholders (e.g., teachers, researchers). Examples of such actions are the identification of problematic items (items that need to be removed, added, or changed), a change to the structure of knowledge components (including changes to the user interface), and the identification of learner groups that need specific treatment.

Typical examples of such applications are described in papers on “closing the loop” (Koedinger et al., 2013; Liu et al., 2014; Cen et al., 2007; Koedinger and McLaughlin, 2016) – an analysis of model results is used to redesign an educational tool and the new version of the tool is evaluated to measure the impact of the change. Another type of actionable insight is reasoning about system features or educational strategies based on the interpretation of model parameters, e.g., the evaluation of help (Beck et al., 2008a) or scaffolding Sao Pedro et al. (2013a). A less typical example of actionable insight is reported by Streeter (2015) – the use of mixture modeling led to the identification of a group of learners with a specific tool setting (disabled sound).

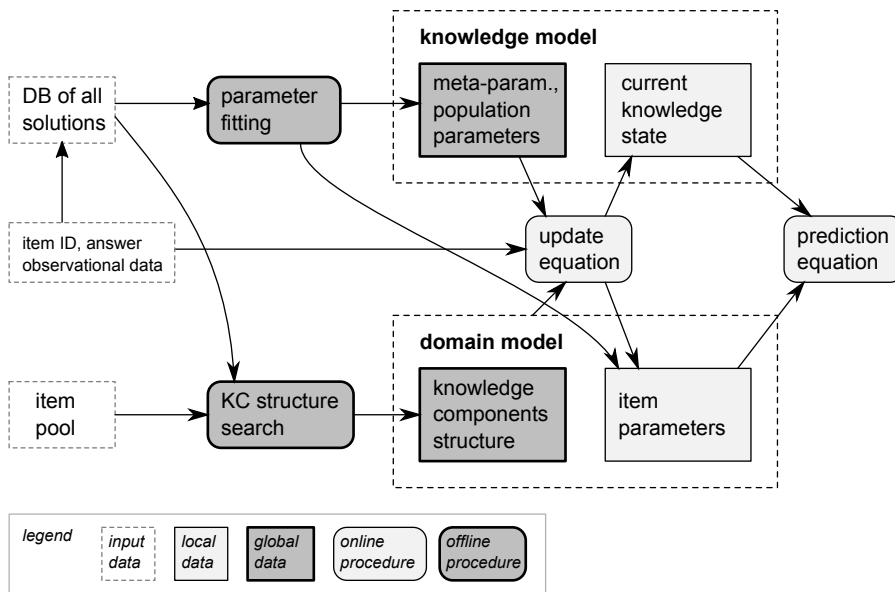


Fig. 2 Learner modeling components. Note that this figure is a zoom of the middle part of Fig. 1.

Another closely related use of learner models (more research oriented) is “discovery with models” (Baker and Yacef, 2009) – a process where an existing model is used as a key component in a new analysis. An example of such an analysis is the use of knowledge estimates in a carelessness detector (Hershkovitz et al., 2013). Insight obtained with the use of learner models can be useful for development not just of interactive educational systems, but also for other types of instructional material (Aleven and Koedinger, 2013).

4 Model Components

After clarifying the overall context of learner modeling, we now discuss learner modeling itself. In this section we describe the main components of a learner model and in the next section we discuss choices for their realization.

Fig. 2 provides an overview of the main components of a typical “long-term” model of learner knowledge. The figure distinguishes between the “data” part of models (parameters, relations) and “procedures” used to compute the data.

4.1 Data

Data for both a knowledge model and a domain model can be further separated into “local data” (information about individual learners and items)

and “global data” (information about the whole population and the domain). Local data are dynamic and they are updated often (after each answer or at least periodically), whereas global data are typically rather static with updates performed either manually or by computationally intensive procedures.

For a domain model, global data consist of the definition of KCs and their structure. A typical example is a mapping of items to KCs and the definition of relations between KCs, e.g., prerequisite relations. Local data consist of the parameters of individual items. A typical example is the difficulty of items, other potential parameters are the easiness of learning or the rate of forgetting.

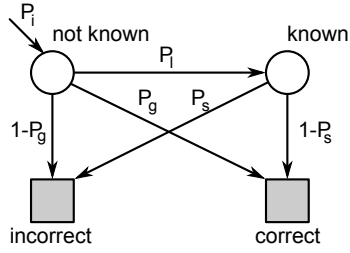
For a knowledge model, global data consist of global parameters for the whole population. A typical example is the distribution of the expected prior knowledge in the population. Global data may also include a definition of learner clusters and parameters for these clusters. Local data contain information about the current knowledge state of individual learners – for each knowledge component there is an estimate of a learner’s knowledge.

4.2 Procedures

The model data are set and updated by specific procedures. The main types of procedures are:

- *KC structure search*. Based on the input of a domain expert or on the analysis of existing data (or combination of both) we search for a “good” specification of KCs and their relation.
- *Parameter fitting*. We use the collected data to optimize model parameters (typically global population parameters and local item parameters).
- *Update equation*. Based on the recent data concerning learner performance, local data are updated, particularly the current knowledge estimate for a learner and often item parameters as well.
- *Prediction equation*. For a given learner and an item the equation predicts future performance. This computation is typically triggered by a request from an instructional policy.

Fig. 2 shows the typical relations between different types of data and procedures. From a practical perspective there is a key difference in speed requirements on different kinds of procedures – this is stressed in the figure by distinguishing between “online” and “offline” procedures. Online procedures are performed as soon as each answer is given. Therefore, for any realistic practical application they have to be very fast (constant or nearly constant computational complexity). On the other hand, a KC structure search and model fitting need only be done offline from time to time and thus they can be more computationally demanding. Nevertheless, due to the volume of data from real applications, even offline procedures should have linear complexity at most; for large data sets it may be preferable to require only a single pass of the data. These computational requirements with differences for individual procedures are sometimes not taken into account in the research literature.



update equation:

$$\text{if } c = 1: \theta' := \frac{\theta(1-P_s)}{\theta(1-P_s)+(1-\theta)P_g}$$

$$\text{if } c = 0: \theta' := \frac{\theta P_s}{\theta P_s+(1-\theta)(1-P_g)}$$

$$\theta := \theta' + (1 - \theta')P_l$$

prediction equation:

$$P_{\text{correct}} = P_g \cdot \theta + (1 - P_s) \cdot (1 - \theta)$$

Fig. 3 The basic structure and equations for the BKT model (c denotes the correctness of an observed answer).

An example of this would be methods based on techniques, such as Markov chain Monte Carlo, which are slow even in the offline analysis and can not be directly used for updating or predicting.

4.3 Examples

Fig. 2 is, of course, a simplification and does not capture perfectly every learner modeling approach. Nevertheless, for the mainstream approaches the figure provides a reasonable fit. As a concrete illustration of the described model components we now discuss two commonly used types of models: Bayesian knowledge tracing and Performance factor analysis. Here we outline only basic versions of these models; extensions are discussed in the next section.

4.3.1 Bayesian Knowledge Tracing

Bayesian knowledge tracing (BKT) (Corbett and Anderson, 1994) is a special case of a hidden Markov model. In BKT, skill is modeled as a binary variable (known/unknown) and learning is modeled by a discrete transition from an unknown to a known state. The basic structure of the model is depicted in Fig. 3; see van de Sande (2013) for a detailed analysis of the model. The basic BKT model uses the following data:

- Global learner data: 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 (a slip), and P_g is the probability of a correct answer when the skill is unlearned (a guess).
- Local learner data: probability θ that a learner is in the known state.
- Global domain data: a definition of knowledge components (sets of items). There are no relations among KCs, i.e., parameters for individual KCs are independent.
- Local domain data: not used in the basic model; extensions of BKT contain such parameters as item difficulties (Pardos and Heffernan, 2011).

The update equation and the prediction equation are shown in Fig. 3 – the probability of being in the known state is updated using a Bayes rule based on an observed answer.

Parameter fitting for the global learner parameters (the tuple P_i, P_l, P_s, P_g) is typically done using the standard expectation-maximization algorithm, alternatively using a stochastic gradient descent or exhaustive search. The specification of KC is typically done manually, potentially using an analysis of learning curves.

4.3.2 Performance Factor Analysis

Performance factor analysis (PFA) (Pavlik et al., 2009) is a specific model from a large class of models based on a logistic function. A common feature of these models is that the data about learner performance are used to compute a skill estimate and this estimate is then transformed using a logistic function into the estimate of the probability of a correct answer. The PFA model specifically uses the following data:

- Global learner data: parameters γ_k, δ_k specifying the change of skill associated with correct and wrong answers for a given KC k .
- Local learner data: a skill estimate θ_k for each KC k .
- Global domain data: a KC difficulty parameter β_k , a Q-matrix Q specifying item-KC mapping; $Q_{ik} \in \{0, 1\}$ denotes whether an item i belongs to KC k .
- Local domain data: not used in the basic model; the model can be easily extended to include item difficulty parameters, see for example the Elo rating system (Pelánek, 2016).

The online equations take the following form (c is the correctness of an answer, i is the index of an item):

- Update equation: $\theta_k := \theta_k + Q_{ik}(\gamma c + \delta(1 - c))$ for each k .
- Prediction equation: $P_{correct} = 1/(1 + e^{-m})$, where $m = \sum_k Q_{ik}(\beta_k + \theta_k)$.

Note that the original formulation of PFA (Pavlik et al., 2009) uses a slightly different notation (using the number of correct and wrong attempts). We have used a transformed but equivalent formulation to highlight the fit of the model to Fig. 2.

Parameter fitting for parameters β, γ, δ can be done easily using standard logistic regression. The Q-matrix is typically manually specified, but can be also fitted using automated techniques like matrix factorization.

5 Overview of Learner Modeling Techniques

Now we discuss several aspects of modeling: the modeling of learning and forgetting, domain modeling, the type of observational data used, and learner clustering. For each of these aspects we provide an overview of available modeling approaches with pointers to specific techniques. These aspects are to a large degree independent and thus there is a vast number of possible combinations that can be used for practical applications.

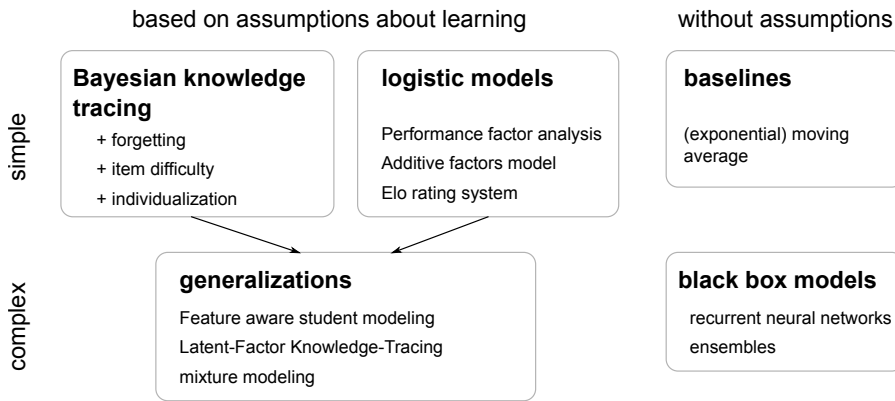


Fig. 4 An overview of basic approaches for modeling learning and forgetting.

5.1 Modeling Learning and Forgetting

The basic goal of a learner model is to estimate the current knowledge and future performance of a learner based on data about their past performance. This can be seen as a typical machine learning task of estimating a hidden state based on noisy observations. A specific feature of learner modeling is the presence of learning and forgetting processes that lead to changes in the hidden knowledge state. The choice of the basic approach to modeling learning and forgetting is thus the key decision in learner modeling. Fig. 4 provides a schematic overview of basic approaches.

Probably the most commonly used approach is the Bayesian knowledge tracing model, which was described in the previous section. From the perspective of learning dynamics, the key assumption of this model is the discrete transition from an unknown to a known state. The basic version of the model presented here has many extensions and variants including forgetting (Khajah et al., 2016), item difficulty (Pardos and Heffernan, 2011), individualization (Pardos and Heffernan, 2010a; Yudelson et al., 2013), and time between attempts (Qiu et al., 2011).

The second major approach to modeling learning is a class of logistic models. In this case skill is modeled by a continuous variable and learning is modeled by a gradual change. Models typically include an item difficulty parameter and the basic principle of predictions is to map a difference between a skill and an item difficulty into the probability of a correct answer using a logistic function $\sigma(x) = 1/(1 + e^{-x})$. Such models are intensively used in item response theory; the basic one parameter logistic model (Rasch model) corresponds to the description provided. Item response theory is typically used in testing and thus it does not consider learning (skill is not expected to change during a test), although extensions that allow for a dynamic change of skill do exist (Wang et al., 2013). In the case of learner modeling a typical logistic model is the Performance factor analysis (Pavlik et al., 2009) described in the previous sec-

tion. Other similar models are the Additive factors model (Käser et al., 2014b; Cen et al., 2006), Instructional factors analysis Chi et al. (2011), and the Elo rating system (Pelánek, 2016). Logistic models are used particularly often in the case of declarative knowledge and the modeling of forgetting (Pavlik and Anderson, 2005; Pelánek, 2015; Rubin et al., 1999; White, 2001; Sense et al., 2016). Logistic models are used not just in the context of education and human learners, but also for the analysis of behavioral experiments with rats (Smith et al., 2004).

The choice between BKT and logistic models is currently not fully resolved. Although some researchers have compared the two approaches (Gong et al., 2010), the results are not conclusive and probably do not generalize – an appropriate choice of a modeling approach depends on a particular domain and on the purpose of a model. However, in many cases researchers pick one of the two modeling approaches without providing any rationale for the choice. The KLI framework may provide guidance and support for this choice. For memory and fluency building processes the knowledge state changes gradually and is thus more naturally modeled by logistic models. BKT assumptions (discrete transition from unknown to known state) are more appropriate for modeling understanding and sense making processes but only for fine grained knowledge components.

Recently, researchers proposed several generalizations and combinations of logistic models and knowledge tracing (Khajah et al., 2014a; González-Brenes et al., 2014; Khajah et al., 2014b). Another flexible modeling approach that generalizes both basic types of models is mixture modeling (Streeter, 2015). Dynamic item response theory models combine modeling based on logistic functions with Bayesian techniques (Wang et al., 2013).

BKT and logistic models differ in their basic assumptions about learning. An alternative approach is to avoid making any specific assumptions about learning. This can be done by using simple approaches like computing a moving average of answer correctness. These simple methods are not based on any specific assumptions about learning, but by discarding or discounting past answers they can model changing skills. A specific simple, yet useful technique is the exponential moving average, where past attempts are weighted by an exponentially decreasing function. Such simple techniques often provide reasonable predictions (Pelánek, 2014; Wauters et al., 2012) and have pragmatic advantages such as the ease of application and computational efficiency. Moreover, the absence of specific assumptions about learning may be an advantage in some circumstances, e.g., a smaller impact of misspecified knowledge components.

We can also compensate for the lack of built-in assumptions about learning by learning patterns from data using more complex machine learning techniques, which include collaborative filtering techniques (Toscher and Jahrer, 2010), ensembles of models (Pardos et al., 2012a), and recurrent neural networks (Piech et al., 2015; Khajah et al., 2016). These models are often able to achieve good predictive accuracy, but at the price of poor interpretability. Since in educational applications we are often concerned with the inter-

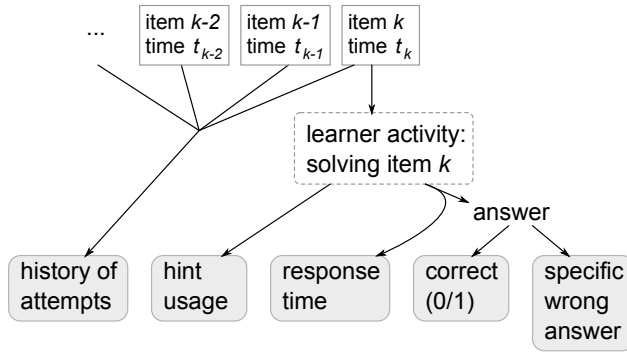


Fig. 5 Basic types of observational data that can be used for modeling knowledge.

pretability and the validity of models, these approaches are not typically used in practical applications at this time.

5.2 Observational Data

Learner knowledge is a latent construct that we try to quantify based on available observational data. The basic information used for learner modeling is the correctness of answers. In many cases this is the only source of data in current use. However, there are many other potentially useful sources of information (Fig. 5): response times, the use of hints, specific values of wrong answers, or the history of previous attempts (including their timestamp).

Response times have been studied thoroughly in the context of item response theory (Goldhammer, 2015). Van Der Linden (2009) provides an overview of conceptual issues in response time modeling in IRT, including a discussion of approaches that model the accuracy and the speed of test takers separately (linked within a hierarchical model). In learner modeling, response times have been typically used in a simpler fashion by combining correctness and response time into a single performance measure (Klinkenberg et al., 2011; Řihák, 2015). Response times have also been used for scheduling the learning of declarative knowledge (Mettler et al., 2011) or in an analysis of slip and guess behavior (Baker et al., 2008). Some modeling approaches even focus specifically on response times (Pelánek and Jarušek, 2015). The current research clearly demonstrates that response times are potentially a useful source of information, but precisely how it is to be exploited awaits further research.

When the interaction of a learner and an item is iterative, such as solving a multistep mathematics problem, we can use data on the solution process. It is not uncommon for educational systems to provide learners with hints. As with response times, data about hint usage provide an indirect but potentially useful source of information about a learner's knowledge. The basic application of hint usage is the use of a 'partial credit' (Wang and Heffernan, 2013; Van Inwegen et al., 2015), i.e., instead of treating the correctness of an

answer as a binary variable, the answer is graded by a partial credit in the interval $[0, 1]$ based on the use of hints and potentially on the response time and other information. Data about hint usage have also been used for modeling help utility (Beck et al., 2008a) and for analyzing hint usage and hint seeking behavior (Goldin et al., 2013). As with response times, the use of hints is potentially useful information. Moreover, the partial credit approach can be combined with most learner modeling approaches. An open issue is how to specify the credit function in a systematic way.

The analysis of wrong answers typically shows that the distribution of mistakes is highly uneven with few common wrong answers (Pelánek and Řihák, 2016). Such common wrong answers can be used for improving model predictions (Řihák and Pelánek, 2016), the labeling of errors (McTavish and Larusson, 2014; Straatemeier, 2014), clustering learners (Merceron and Yacef, 2005), and affect detection (Wang et al., 2015). The use of specific wrong answers in learner modeling may be useful particularly in conjunction with an explicit modeling of misconceptions in a learner model (Liu et al., 2016). The research analyzing wrong answers is mostly very recent and provides mainly illustrations of the potential of the data source: more research in this direction is needed.

Another potential source of information is the history of attempts – what a learner did before submitting a particular answer. This source of data can serve many different purposes. Models can, for example, incorporate sequential (ordering) effects (Pardos and Heffernan, 2009; Tang et al., 2015), an effect of a new session (Qiu et al., 2011), or a contextual estimation of slip and guess parameters (Baker et al., 2008). A specific (but common) case of sequential effects is the case where items are presented in a fixed order with items of increasing difficulty. In this case modeling often faces an identifiability problem – it is hard to distinguish learning from increase in item difficulties (González-Brenes et al., 2014; Khajah et al., 2014a; Pelánek and Jarušek, 2015).

The history of attempts may contain information not just about interactive items, but also about instructional steps, e.g., information that a learner viewed a video or went through lecture materials. Such information can be included into models to improve model predictions, e.g., Performance factor analysis has been extended in this way into Instructional factors analysis (Chi et al., 2011), the BKT model has been extended to include scaffolding and tutor context (Sao Pedro et al., 2013a). Such models may be used for evaluating the quality of instructional materials (MacHardy and Pardos, 2015).

The importance of different types of observational data depends on the type of relevant knowledge components and learning processes. For example, response times are particularly relevant for fluency building, whereas the analysis of hint usage and the modeling of misconceptions are relevant only for understanding and sense-making processes.

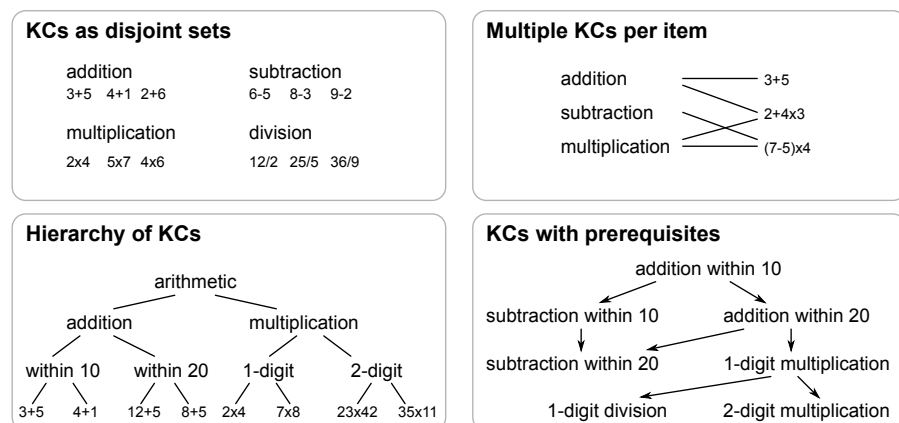


Fig. 6 Domain modeling – the illustration of basic approaches.

5.3 Domain Modeling

Domain modeling is concerned with the assignment of individual items to knowledge components and with the modeling of relations among KCs. The basic approaches to domain modeling are illustrated in Fig. 6. The simplest approach is to consider KCs as disjoint sets of items. This can be extended in three main directions: multiple KCs per item, a hierarchy of KCs (capturing skills of different granularity), and relations between KCs (particularly their prerequisite structure).

A principled way of modeling KCs with relations is provided by (Dynamic) Bayesian networks (Millán et al., 2010; Conati et al., 2002; Käser et al., 2014a; Carmona et al., 2005; Käser et al., 2013a). Bayesian networks can model not only skill relations, but also the uncertainty of estimated skill parameters. However, the use of Bayesian techniques makes high demands on computational resources, because parameter estimation becomes difficult. Another approach to modeling skill relations based on formal foundations is knowledge space theory (Doignon and Falmagne, 2012) and its variants (Desmarais et al., 2006). This approach is useful particularly for modeling prerequisite relations. For practical applications it is useful to consider more heuristic approaches, e.g., a hierarchical extension of the Elo rating system (Niznan et al., 2015b).

A rather difficult problem for learner modeling is posed by the presence of multiple skills per item, i.e., solving an item requires knowledge spanning multiple KCs. An item-skill mapping is in this case a bipartite graph (illustrated in Fig. 6), typically coded using a so called Q-matrix (Tatsuoka, 1983; Barnes, 2005; Desmarais, 2011). The use of multiple skills per item leads to the “credit assignment problem”: How do we model the relations between performance and related skills? If the learner answers incorrectly, which skill is to be “blamed” for this mistake? Researchers have explored many ways of combining several skills, e.g., using a compensatory (additive) model (Ayers

and Junker, 2006), conjunctive (product) model (Cen et al., 2008; Koedinger et al., 2011; Beck et al., 2008b), logistic regression (Xu and Mostow, 2012), or taking the weakest skill (Gong et al., 2010). A related set of models (used more commonly in psychometrics) is the NIDA/NIDO/DINA/DINO family of models, e.g., NIDA model (noisy input, deterministic and) (Junker and Sijtsma, 2001) or DINA (deterministic inputs, noisy and) model (De La Torre, 2009). It seems that there is no universal solution to the credit assignment problem; a suitable approach depends on a particular domain and the type of knowledge component. Moreover, in many cases parameter estimation is computationally demanding and for practical application this also needs to be taken into account.

Once we decide which domain modeling approach to use, we have to find a specific domain model (e.g., the mapping of items to KCs, relations between KCs). As the illustration in Fig. 6 shows, even for a simple domain like basic arithmetic it is far from clear what makes a good domain model. What should be the granularity of KCs? What relations should be modeled? What is the assignment of items to KCs? The model can be provided either by a domain expert or determined using skill discovery. An example of a domain model of basic arithmetic specified by an expert is provided by Käser et al. (2013a).

Since the manual specification of a domain model is time consuming and error-prone, automatic model discovery has received significant attention. Researchers have studied many approaches including learning factor analysis using A* search (Cen et al., 2006), matrix factorization (Thai-Nghe et al., 2011; Desmarais, 2011; Lan et al., 2014), Chinese restaurant process (Lindsey et al., 2014), spectral clustering (Boroš et al., 2013), or analysis using simulated students (Li et al., 2011). Specific techniques focus on discovering the prerequisite structure (Chen et al., 2015; Scheines et al., 2014; Gasparetti et al., 2015; Chen et al., 2016).

It is, of course, possible to combine input from experts with automated techniques based on data. As a basic step, we can compare several manually created domain models and choose the one that fits the data best. A specific example of such an approach is the comparison of models of different skill granularity (Feng et al., 2006; Pardos et al., 2010; Koedinger et al., 2016). A more sophisticated approach is to take an expert provided model and refine it based on data (Desmarais and Naceur, 2013; Desmarais et al., 2014; Nižnan et al., 2014).

When choosing a domain modeling approach for a particular application, it is useful to take into account the KLI framework. For understanding and sense-making processes it is typically important to take into account a prerequisite structure of the domain. On the other hand, for memory and fluency-building processes it is typically sufficient (and preferable) to use a simpler domain model, such as “KCs as disjoint sets”.

5.4 Learner Clustering and Individualization

The basic learner modeling approach is to treat all learners as coming from the same homogeneous population. This is clearly a simplifying assumption as a learner population is never completely homogeneous. The important question is whether learners fall into sufficiently different clusters, which could be exploited for learner modeling. Such broad diversity may appear when systems are used by both children and adult learners, in the presence of learners with different forms of learning disabilities, and among learners from different countries.

The clustering of learners (Hämäläinen et al., 2014) has been explored by many researchers, e.g., in mathematics learning with respect to different forms of dyscalculia (Käser et al., 2013b; Carlson et al., 2013) and for the evaluation of logic learning (Merceron and Yacef, 2005). Basic clustering is done using individual answers, but it is also possible to cluster sequences (Desmarais and Lemieux, 2013; Klingler et al., 2016). A principled approach to modeling learning in the presence of learner clusters is mixture modeling (Streeter, 2015), which reveals the clusters and fits their learning patterns at the same time (using the EM algorithm). A specific type of cluster is “wheel-spinning learners” (Beck and Gong, 2013). These are learners who are unable to master a topic, as happens for example, when learners do not have the prerequisite knowledge.

Detected learner clusters can be applied to improve predictions by using different models or parameter values for each cluster (Trivedi et al., 2011; Pardos et al., 2012b; Gong et al., 2012) and for gaining insight into learner behavior, e.g., identifying user interface issues (Streeter, 2015).

A closely related issue is the level of granularity and individualization in a learner model. Some model parameters can be considered either as global (population level) or individual. The impact of this choice has been studied particularly in the context of the BKT model and parameters for prior knowledge and speed of learning (Pardos and Heffernan, 2010a; Lee and Brunskill, 2012; Yudelson et al., 2013; Pardos and Xu, 2016). Individualization increases the number of model parameters and carries the risk of overfitting. The individualized parameters are fitted using only few data points and thus can be significantly influenced by the noise in data. To avoid this risk, it is also possible to consider “individualization” on the level of learner groups, e.g., having the same parameters for all learners in one class (Wang and Beck, 2013) or in a learner cluster automatically detected by one of the above-mentioned techniques.

The modeling of learner populations is closely associated with domain modeling, particularly to the granularity of KCs. For coarse grained KCs (e.g., “fractions”, “capitalization rules”) the heterogeneity of learner population can be an important factor as different types of learners may be strong in different parts of a coarse KC. For fine grained KCs clustering should not be a very important factor – it should be sufficient to differentiate learners by their skill estimates.

5.5 Summary of Modeling Aspects

The overview provided in this section shows that there are many different aspects of learner modeling and many possible choices for each aspect. Where should we focus our modeling effort for a particular application? Fig. 7 provides a summary of the main modeling aspects and their importance with respect to model purposes and with respect to learning processes (as defined in KLI).

The figure provides only a basic orientation. The “importance” of modeling aspects is not precisely defined and at the current stage of research in many cases it may be disputable. In particular, the relevance of modeling techniques for different learning processes has not yet been thoroughly studied and the proposed mapping should be verified and clarified in future research.

The main point of the figure is that we cannot reach any universal conclusions about questions like “Which learner model is better?” or “Is it useful to use X in learner modeling?” (for some source X of data about learners such as response times or hint use). Answers to such questions depend on the context in which learner modeling happens.

For illustration consider two specific contrasting examples. At first, let us consider the use of learner modeling for a personalized practice of foreign language vocabulary. A model is used for automatic item choice (a word to practice) with the goal of supporting memory and fluency processes. In this case the modeling approach should clearly focus on learning and forgetting processes; probably using one of the models from the family of logistic models. Timing information is here clearly important: the period since the previous exposure to an item in the assessment of forgetting, the response time in the assessment of fluency. On the other hand, the role of modeling KCs structure is not fundamental.

Secondly, let us consider the use of learner modeling for the analysis of historical data about the usage of a fraction tutor that is concerned with understanding and sense-making processes. Moreover, let us assume that the purpose of a model is getting actionable insight that will be used for a manual redesign of the tutor, e.g., the evaluation of the available hints and scaffolding problems, or an analysis of a suitable granularity of the knowledge components used. In this case some version of the BKT model should be relevant and domain modeling is clearly very important. From observational data it may be useful to utilize data about the use of hints and other instructional materials. Timing information is not as important as in the vocabulary case.

Another important issue is the combination of different aspects. As noted at the beginning of this sections, the aspects described are to a large degree independent, e.g., the approaches depicted in Fig. 4, Fig. 5, and Fig. 6 can be combined in many ways. Nevertheless, some combinations are more difficult than others. For example, it is easier to use “multiple KCs per item” with logistic models than with BKT. To add new observational data into simple models of learning, it is necessary to hand-craft the model features and their role in a model, whereas with neural networks the addition of new observa-

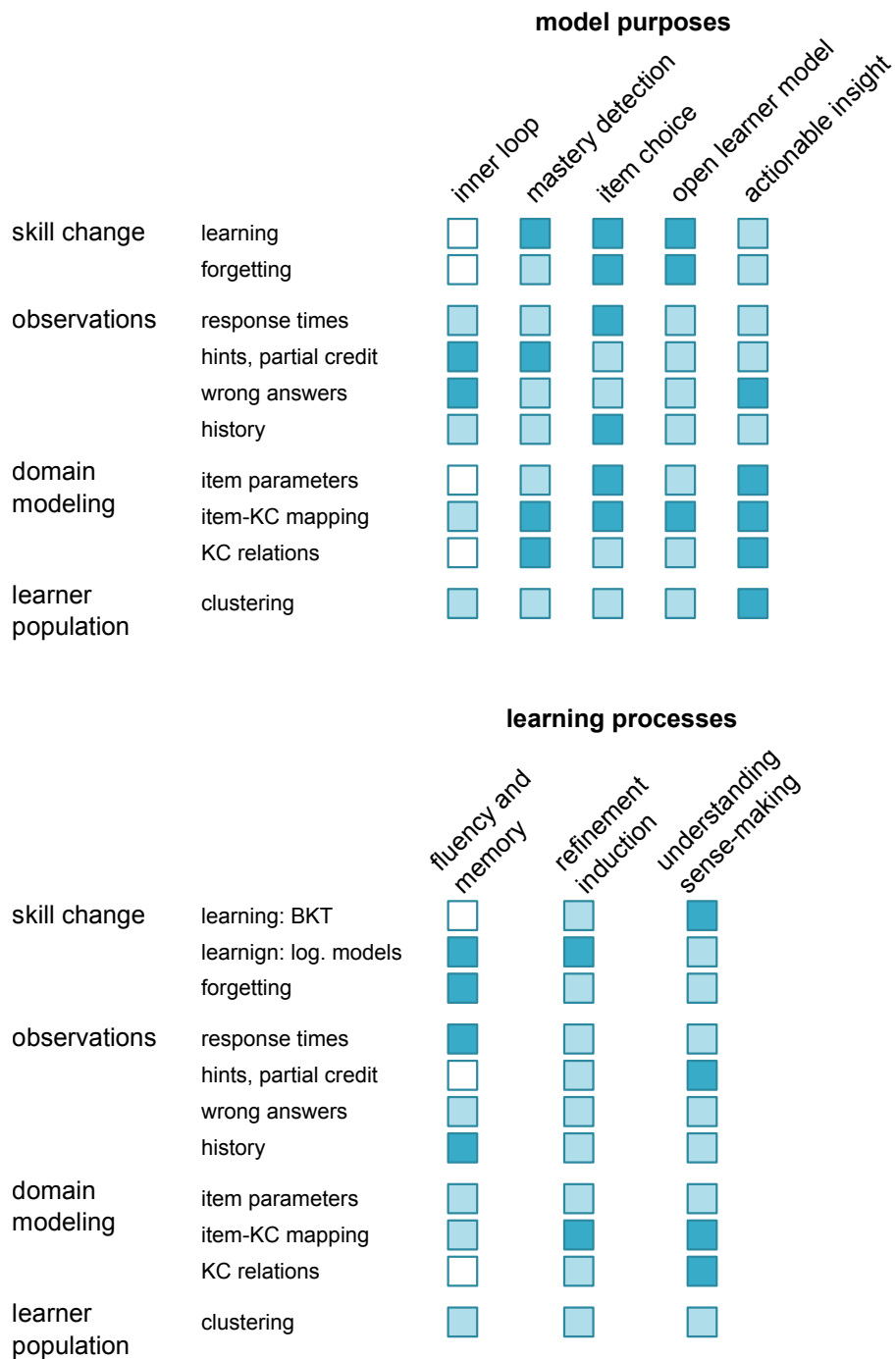


Fig. 7 The relevance of model aspects with respect to model purpose and learning processes (KLI). Darker color denotes higher importance.

tional data may be trivial even though it may require significantly more data and longer processing times.

We hypothesize that for learner modeling to progress, it is more important to clarify the mapping between modeling methods and contexts (as outlined in Fig. 7) and to explore the interoperability of all the aspects of modelling than to keep making incremental innovations in individual modeling methods.

6 Model Fitting and Evaluation

Once specific learner modeling approach has been selected, we still need to find values of model parameters. In this step we also face several choices. What procedure do we use for parameter filtering? What criteria do we use to choose among competing models or parametrizations of a chosen model? How do we address overfitting? How do we treat potential biases in data and in parameter estimation? As was the case in choosing a modeling approach, there are no universally applicable answers to these questions. An appropriate choice of model evaluation methodology depends on the specific domain, the type of knowledge components, and especially the purpose of the model. We will now discuss the most important methodological issues in model evaluation and highlight their relation to the context of modeling.

6.1 Metrics for Model Comparison

To compare models we need to quantify their quality. This is typically done by comparing their predictive accuracy, i.e., their ability to predict future learner performance. Table 3 provides examples of several commonly used metrics for quantifying predictive accuracy; see Pelánek (2015) for definitions and properties of individual metrics. These metrics measure model performance using “within system” predictions. An alternative approach is to use external measures like the correlation of knowledge estimated by the model with an external test, e.g., a high stakes test outside of a system.

The choice of a specific metric can have significant impact on model comparison and parameter fitting (Pelánek, 2015; Huang et al., 2015b; Stamper et al., 2013). Model comparisons can be influenced even by details of metric computation that are typically not explicitly described in research papers. For example, Khajah et al. (2016) discuss the issue of averaging in the AUC computation: using all predictions to compute AUC versus computing AUC on a per-skill basis and then taking an average. They found that it significantly inflated differences between models in a previous work.

The appropriate choice of a metric depends on the purpose of the model to be evaluated. For example if the model output is used by an instructional policy that considers absolute values of predictions, it is not meaningful to perform model evaluation using the AUC metric, since it considers only the relative ordering of predictions (Pelánek, 2015).

Table 3 Examples of metrics for the predictive accuracy of learner models.

type of metric	specific metric	
probabilistic understanding of errors	RMSE	root mean square error
	LL	log-likelihood
	R^2	normalized RMSE
qualitative understanding of errors	Acc	accuracy
	F1	harmonic mean of precision and recall
assessing ranking of examples	AUC	area under the ROC curve

Comparing predictive accuracy is important even in cases where the primary purpose of a model is not predictions (e.g., in “discovery with models”), since parameter fitting procedures typically optimize predictive accuracy – often implicitly log-likelihood. For this reason, performance metrics are quite central to learner modeling. But the way we interpret the results is not universal and depends heavily on the specific purpose of a model. A particular recurring issue in learner modeling research is the importance of small differences in predictive accuracy. Research papers often report improvement in learner modeling with statistically significant, but rather small differences in predictive accuracy. Some researchers have expressed doubts about the usefulness of such improvements (Beck and Xiong, 2013). In many cases, the impact of a slightly improved predictive accuracy is negligible, particularly if the model is used only to provide predictions and the predictions of competing models are highly correlated. However, when models are used for actionable insight, even a small difference in their metric value results in a significant impact on learner practice (Yudelson and Koedinger, 2013; Pardos et al., 2012c; Liu et al., 2014).

One of the common purposes of learner models is to provide input data for an instructional policy. In these cases, we can evaluate the impact of different models on the decisions made by the policy. So far, this type of model evaluation has been done mainly for the mastery learning policy by measuring the impact of a model on the number of learner practice opportunities (Lee and Brunskill, 2012; González-Brenes and Huang, 2015; Rollinson and Brunskill, 2015; Käser et al., 2016). Most evaluation techniques used in this research are, however, closely intertwined with Bayesian knowledge tracing – they incorporate assumptions of this modeling approach and are thus applicable only to models based on BKT.

To better understand model predictions, it is useful to go beyond a single number provided by accuracy metrics. Specifically, it is useful to analyze the reliability and the resolution of predictions, which can provide useful insight into model behavior and provide directions for model improvement (Pelánek, 2015). It may be also useful to check for specific undesired behaviors. For example, some models can predict mastery even for a learner who always answers incorrectly – for example a basic mixture model described by Streeter (2015). If such cases are uncommon in the testing data, the impact on an

accuracy metric is negligible, yet for practical applications such model behavior can have a significant undesirable impact by undermining users' trust in the system.

6.2 Parameter Fitting and Analysis

For a specific model there may be several applicable parameter estimation procedures and the choice of a parameter fitting procedure may influence the resulting parameters and the predictive accuracy of a model. There is a particularly rich literature for estimating the parameters of the BKT model, e.g., Pardos and Heffernan (2010b); Hawkins et al. (2014); Beck and Chang (2007); Falakmasir et al. (2013). This issue was also studied for logistic models, e.g., Gong et al. (2010); Papoušek et al. (2014). The reported results suggest that different parameter estimation procedures often lead to similar parameters (and consequently also similar predictive accuracy), but there are often significant differences in the computational complexity of procedures.

The issue of computational efficiency is very important in practice, but sometimes does not get the research attention it deserves – in particular the distinction between updates that need to be performed online versus offline, as discussed in Section 4. Complex learner models may require computationally demanding parameter fitting procedures, such as models based on Bayesian networks which are typically fitted by Markov chain Monte Carlo (Khajah et al., 2014a). Such procedures may not be applicable for online educational systems or even for the offline analysis of large data sets, e.g., Khajah et al. (2014a) reports a runtime in excess of 10 minutes for a data set with 110,000 attempts, which is quite small in terms of many practical applications.

What is the relative importance of the computational efficiency of parameter fitting on one hand and the predictive accuracy of the fitted model on the other hand? The answer depends on the purpose of the model. If the model is used only to provide predictions used by an instructional policy to make online decisions, computational efficiency is a key factor whereas small differences in predictions are not very significant. If, on the other hand, we use a model for obtaining “actionable insight” that is interpreted “offline” by humans, small differences in predictive accuracy may be very important since they may lead to different conclusions from the discovery process Pardos et al. (2012c). The computational efficiency of parameter fitting is in this case less important as long as it scales to the size of a particular data set.

If the primary purpose of a model is open learner modeling or getting actionable insight, we are more interested in model parameters than in model predictions. In these cases, previous research has sometimes been based on the implicit argument, “if prediction accuracy is improved then the additional factor in the model is meaningful and parameters can be interpreted”. For examples see Gong and Beck (2011); Beck et al. (2008a); Huang et al. (2015a).

This approach, however, is not sufficient. Before seriously interpreting parameter values, it is necessary to analyze their consistency – parameters may

have different values when the model is trained on different training sets or when different parameter fitting procedures are used. There are several reasons for this including noise in the data, model identifiability issues, and local optima in parameter fitting. The analysis of parameter consistency has been recently proposed as one aspect in a framework for multifaceted evaluation of models (Huang et al., 2015b). Different approaches can be used to perform such analyses, e.g., estimating confidence intervals using bootstrapping, analyzing parameter values for independent subsets of data (Pelánek and Jarušek, 2015), analyzing parameters for closely similar items (Klinkenberg et al., 2011; Arroyo et al., 2010), or analyzing the correlation of parameter values with some external measure like a pretest scores (Gong et al., 2011).

6.3 Cross-validation Approach

An important general issue in machine learning is overfitting: the aim of machine learning is to develop models that do not just fit training data, but also generalize to new circumstances. One approach to avoiding overfitting is to use performance metrics like Akaike information criterion (AIC) or Bayesian information criterion (BIC), which are extensions of log-likelihood penalizing model complexity. Another approach is to use cross-validation, which involves evaluating a model on a separate test set. Although under certain circumstances these approaches may be asymptotically equivalent (Stone, 1977), practical applications typically exhibit far from asymptotic behavior. More importantly, cross-validation allows us to take into account the type of generalization that is relevant for a particular model purpose. Therefore, cross-validation is a better approach for model evaluation in most applications. A possible reason to use metrics like AIC and BIC is better computational efficiency, which is important in computationally demanding “search for a model” procedures like learning factors analysis (Cen et al., 2006). In such cases, it is important to analyze relations between these metrics and cross-validation results, as done for example by Stamper et al. (2013).

Cross-validation evaluates model ability to generalize by using different data for training (parameter fitting) and testing (measuring predictive accuracy). For educational data it is necessary to pay close attention to the way data are partitioned into a training set and a test set. In many areas a division of data points into training and testing set can be done by simple random selection. In the case of learner modeling simple random allocation is incorrect – since we deal with sequential data, we would end up using future actions for predicting past actions. We also need to take into account the asymmetry between learners and items. Items are usually rather fixed, whereas new learners arrive continuously, so we mainly want to be able to evaluate generalizations across learners actions.

Fig. 8 illustrates several basic approaches to performing cross-validation. The figure uses two basic dimensions:

- The division of data with respect to learners:

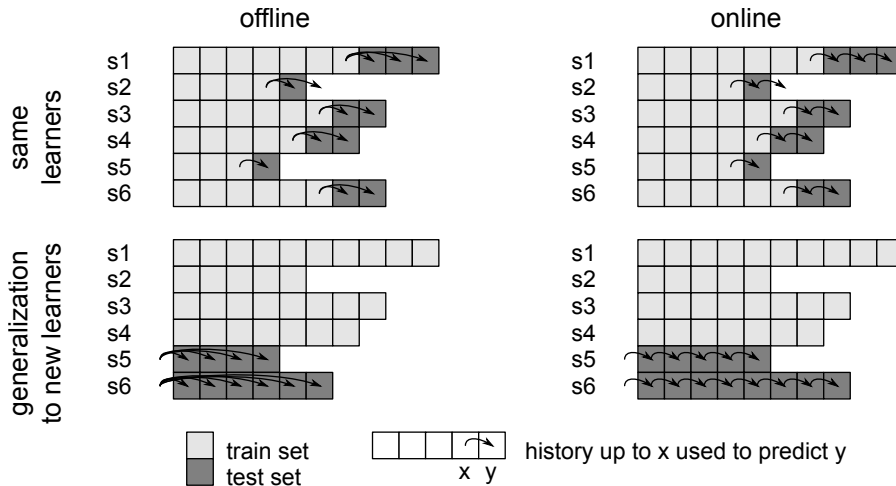


Fig. 8 An illustration of the basic options for cross-validation methodology. Similar way to illustrate division into training and testing set was previously used by Khajah et al. (2014a) and Reddy et al. (2016).

- *Testing on the same learners.* The beginning of each sequence is in a training set, the end of the sequence is in a test set.
- *Testing on new learners.* All attempts of a learner are either in a training set or in a test set.
- The update of predictions:
 - *Offline evaluation.* Predictions for a whole learner’s sequence are made at the same time.
 - *Online evaluation.* Predictions are continuously updated after observing each answer.

To illustrate the scope of different cross-validation approaches, we provide examples of specific methods used in previous work. Online evaluation on new learners was used, for example, by Streeter (2015); Nižnan et al. (2015b); Pardos and Heffernan (2011); it is also typically employed in research based on BKT. Offline evaluation on new learners was used, for example, by Käser et al. (2014b); Klingler et al. (2015); and also in other works using the AFM model. González-Brenes et al. (2014) used offline evaluation on new learners, but only the second half of sequence is used for evaluation. Offline evaluation on the same learners was used for example by Khajah et al. (2014b); Pelánek and Jarušek (2015) (using the last 20% of attempts). KDD cup 2010 data set (used for example by Thai-Nghe et al. (2011)) has the same learners in a test set with a combination of online and offline evaluation – for each learner predictions are made for a single problem that involves multiple steps. Finally, some researchers use only a single attempt per learner for the evaluation of predictions (Pardos and Heffernan, 2010a; Reddy et al., 2016).

A proper cross-validation methodology depends on the specific purpose of the model being evaluated. In the standard scenario where the purpose of

the model concerns “predictions”, the preferable cross-validation methodology should be “online, generalization to new learners”, because it directly corresponds to the actual application of a learner model in an educational system.

So far we have discussed learner-stratified cross-validation, which tests generalization to new learners. Depending on a particular application and the purpose of a model, we may also need to explore other types of generalization: generalization to new items in the case of knowledge components with a nontrivial churn rate, such as those in educational systems for information technology, generalization to new knowledge components, or generalization to new learner populations, which is relevant particularly in the case of discovery with models. For these situations, it is necessary for data to be adequately divided into a training set and a test set – previous research has shown that different kinds of stratification may lead to different results (Sao Pedro et al., 2013b).

6.4 Data Collection

Data collection mechanisms are another aspect where the context of learner modeling is important. This is especially the case when using naturally occurring observational data, which is a common practice in research on learner modeling. The way in which this data were collected may significantly influence the results such as parameter values or the comparison of models (Pelánek et al., 2016). Several aspects of data collection exhibit this influence.

A common feature present in many educational data sets is attrition bias – a selection bias caused by differences in the way learners use an educational system. A typical example is mastery attrition bias caused by the mastery learning principle explicitly implemented in a system (Nixon et al., 2013). Attrition bias can be also caused by self-selection (Papoušek et al., 2016) and it can have a significant impact on learning curves and fitted learner parameters (Käser et al., 2014b; Murray et al., 2013).

Another important aspect of data collection is the ordering of items. If all learners attempt items in a similar order, it may be impossible to disentangle learning from an increase in problem difficulty. This confounding effect has been noted in different forms in several recent works (González-Brenes et al., 2014; Khajah et al., 2014a; Pelánek and Jarušek, 2015).

This effect disappears once we use an adaptive choice of items and the ordering is personalized for each learner. This, however, creates another, potentially more complex problem for model evaluation – it creates a feedback loop between learner modeling and data collection (Nižnan et al., 2015a; Pelánek et al., 2016). For example, if data are collected using an adaptive system that provides learners with items of appropriate difficulty, even a simple baseline model achieves a good predictive accuracy and differences between models may become small even though the consequences of using different models in an application would be large (Pelánek et al., 2016).

The most important step in overcoming biases caused by data collection is to be aware of them. We need to explicitly formulate them and to consider their importance with respect to a particular purpose of a studied learner model. If we have control over data collection, it may be useful to introduce controlled randomization into data collection, as used for example by Papoušek et al. (2016). If we only have access to historical data sets, it is useful to filter the available data to test the robustness of results, for example, by limiting the number of answers per learner to reduce the impact of attrition bias.

6.5 Summary of Model Fitting and Evaluation

Model fitting and evaluation entails a wide range of choices (e.g., evaluation metric, parameter fitting procedure, cross-validation approach). The current state-of-the-art does not provide crystal clear guidance to making these choices. The first step is making these choices explicit and connecting them to the specific purpose of a model – currently many choices are made implicitly, often without a proper description or rationale (particularly cross-validation approaches).

To illustrate the importance of the purpose of the model on decisions, we use the same two examples as in the previous section. The first example is the use of learner modeling for a personalized practice of foreign language vocabulary, where a model is used for automatic item choice. This is an online application of a learner model, i.e., it requires an online parameter fitting and online cross-validation methodology. The key model outputs are predictions of learners' performance on specific items. Comparison of models should focus on the predictive accuracy of models and on the impact of models on the final choice of items that are presented to learners. Ultimately, we are concerned with the impact of models on learning, i.e., an ideal evaluation should be done using a randomized control trial using the real system. If the evaluation is done using historical data, it is necessary to pay attention to the potential biases present in data, particularly when the data were collected using a system that already implements adaptive behavior.

The second example is the use of learner modeling for the analysis of historical data about the usage of a fraction tutor, where the purpose of a model is getting actionable insight for manual improvement of the system. For this application the parameter fitting is done offline and thus its speed is not fundamental. We are interested primarily in model parameters, not model predictions, so the analysis should focus primarily on the stability of model parameters. For cross-validation it may be reasonable to use the offline approach, but with more focus on the level of stratification. For the verification of actionable insight, it may be useful to test the generalization to new learner populations, i.e., to use population-stratified cross-validation.

7 Discussion and Future Work

In learner modeling we have to make many decisions. Research literature currently offers an abundance of choices, but it provides little guidance on how to choose a particular approach for a specific situation. To deal successfully with these decisions we need to take a wider context of learner modeling into account. What types of knowledge components and learning processes are relevant? What sources of data can be used for learner modeling? What is the main purpose of learner modeling in a particular application? How will the outputs of learner modeling be used? Some conflicting results in research literature may be due to differences in such contexts. Beel et al. (2016) discuss similar issues for the closely related domain of recommender systems.

In this work we presented an overview of the current state of learner modeling taking this wider context into account. The arguments presented here have several consequences for both developers and researchers.

7.1 Developers' Perspective

The performance of real world educational systems is to a large degree determined by the “weakest link”. For example, it is important to consider all the links in Fig. 1. It is typically better to have a simple implementation of all important components than to have a very sophisticated model of learning, but poor implementation of an instructional policy and missing an open learner model. Previous work already noted this risk: “it is easy to get carried away by the sheer intellectual challenge of assessment and to overbuild this part of the tutoring system” (Vanlehn, 2006). Moreover, real-world machine learning systems often have high maintenance costs (Sculley et al., 2015). From the developers' perspective it is thus preferable to use simple learner models unless there is a clear reason to prefer more complex models.

Even when we decide to employ “simple” models, we still need to make many choices, e.g., about the granularity of the knowledge components used or the scope of observed data. Throughout this paper we have argued that the KLI framework can provide guidance for making these choices. An explicit clarification of the types of knowledge components and learning processes relevant for a particular application is useful not only for choosing a proper learner model, but it can be also helpful in other design decisions (e.g., the organization of the user interface or the design of interactive activities). Similarly, it is important to clarify the purpose of the model for a particular application and to take it into account when choosing a modeling approach. For example, model interpretability is not fundamental for an automatic instructional policy, but very important for getting actionable insight (“human-in-the loop”).

Fig. 7 provides an overview of the main relations between types of learning processes, purposes of models, and aspects of learner modeling. Given the state-of-the-art, the figure can provide only a basic guidance. Reports on practical case studies that explicitly focus on the relations between the KLI

framework and learner modeling decisions would be beneficial for further clarification and verification of the proposed guidelines.

From the developer’s perspective an important issue is model portability – it is advantageous when models can be transferred from one environment to another. The issue of portability has received attention in learner modeling only recently (Valdés Aguirre et al., 2016). Further research in this direction should take the context of learner modeling explicitly into account – it is probable that portability is feasible only in environments that have a very similar context, i.e., having the same purpose of the model and addressing similar types of knowledge components and learning processes.

7.2 Researchers’ Perspective

Putting learner modeling into a relationship with the KLI framework and purposes of models provides an inspiration for future research work. The arguments presented about the relationships outlined in Fig. 7 should be further specified and experimentally tested.

For illustration we present several specific hypotheses of this type. These hypotheses are mostly based on arguments presented in this work and on previous research, but currently there is not sufficient evidence to provide clear backing of them. Moreover, we expect that even if they are valid in their general form, further research will lead to adding clarifications and more nuanced formulations.

- Hypothesis 1: The relative performance of Bayesian knowledge tracing versus logistic models of learning depends on the type of relevant learning processes. Logistic models are better for modeling fluency and memory processes, while Bayesian knowledge tracing is better for understanding and sense making processes.
- Hypothesis 2: The modeling of forgetting is very important for fluency and memory processes, but for understanding and sense making processes it brings only slight improvement (as measured by the predictive accuracy of models or by an impact on an automated instructional policy).
- Hypothesis 3: If the model is used for mastery detection, it is more important what data are used for modeling than the exact details of models, e.g., incorporating response times to mastery criteria has higher impact than using a different model of learning. Slightly different models with the same input data lead to very similar mastery decision.
- Hypothesis 4: If the model is used for actionable insight, then even slightly different models with the same input data and with similar predictive accuracy can lead to different conclusions and thus different actions being taken.

Model evaluations presented in research papers should pay more attention to the purpose of models. The purpose of evaluated models should be explicitly stated and evaluation methodology should be selected accordingly for the given

purpose (e.g., the choice of a metric and a cross-validation approach). From the research perspective an important type of model purpose is “actionable insight”. This type of analysis is typically based on the interpretation of model parameters. Therefore, for this purpose it is important to pay specific attention to the consistency of parameters and not just analyze the prediction accuracy (Huang et al., 2015b). It is also important to consider the potential impact of data collection or at least to describe the way data were collected and to mention limitations due to data collection (Pelánek et al., 2016).

An interesting problem connecting the researchers’ and developers’ perspective is the identification of leverage points (Meadows, 1999) – modeling decisions with the highest impact. It would be useful to develop general techniques and guidelines that would help to identify leverage points for a particular learner modeling application.

Acknowledgments

The author thanks members of the Adaptive Learning group at Masaryk University for valuable discussions and for their work on the development of practical educational system, which inspired many aspects of this overview. The author also thanks anonymous reviewers for their valuable comments, which significantly improved the structure and presentation of the paper.

References

- Aleven V, Koedinger KR (2013) Knowledge component (kc) approaches to learner modeling. In: Sottolare RA, Graesser A, Hu X, Holden H (eds) *Design Recommendations for Intelligent Tutoring Systems*, vol 1, US Army Research Laboratory, pp 165–182
- Aleven V, Sewall J (2016) The frequency of tutor behaviors: A case study. In: *Proc. of Intelligent Tutoring Systems*, Springer, pp 396–401
- Aleven V, McLaren BM, Sewall J, Koedinger KR (2006) The cognitive tutor authoring tools (ctat): preliminary evaluation of efficiency gains. In: *International Conference on Intelligent Tutoring Systems*, Springer, pp 61–70
- Anderson JR, Corbett AT, Koedinger KR, Pelletier R (1995) Cognitive tutors: Lessons learned. *The journal of the learning sciences* 4(2):167–207
- Arroyo I, Mehranian H, Woolf BP (2010) Effort-based tutoring: An empirical approach to intelligent tutoring. In: *Proc. of Educational Data Mining*
- Ayers E, Junker B (2006) Do skills combine additively to predict task difficulty in eighth grade mathematics. In: *Proc. of Educational Data Mining: Papers from the AAAI Workshop*
- Baker RS (2016) Stupid tutoring systems, intelligent humans. *International Journal of Artificial Intelligence in Education* 26(2):600–614
- Baker RS, Yacef K (2009) The state of educational data mining in 2009: A review and future visions. *Journal of Educational Data Mining* 1(1):3–17

- Baker RS, Corbett AT, Aleven V (2008) More accurate student modeling through contextual estimation of slip and guess probabilities in bayesian knowledge tracing. In: Proc. of Intelligent Tutoring Systems, Springer, pp 406–415
- Barnes T (2005) The q-matrix method: Mining student response data for knowledge. In: American Association for Artificial Intelligence 2005 Educational Data Mining Workshop, pp 1–8
- Beck J, Xiong X (2013) Limits to accuracy: how well can we do at student modeling? In: Proc. of Educational Data Mining
- Beck JE, Chang Km (2007) Identifiability: A fundamental problem of student modeling. In: User Modeling, Springer, pp 137–146
- Beck JE, Gong Y (2013) Wheel-spinning: Students who fail to master a skill. In: Proc. of Artificial Intelligence in Education, Springer, pp 431–440
- Beck JE, Chang Km, Mostow J, Corbett A (2008a) Does help help? introducing the bayesian evaluation and assessment methodology. In: Proc. of Intelligent Tutoring Systems, Springer, pp 383–394
- Beck JE, Pardos ZA, Heffernan NT, Ruiz C (2008b) The composition effect: Conjunctive or compensatory? an analysis of multi-skill math questions in its. In: Proc. of Educational Data Mining
- Beel J, Breitingner C, Langer S, Lommatzsch A, Gipp B (2016) Towards reproducibility in recommender-systems research. User Modeling and User-Adapted Interaction 26(1):69–101
- Boroš P, Nižnan J, Pelánek R, Řihák J (2013) Automatic detection of concepts from problem solving times. In: Proc. of Artificial Intelligence in Education, Springer, LNCS, vol 7926, pp 595–598
- Bull S, Kay J (2007) Student models that invite the learner in: The smil:() open learner modelling framework. International Journal of Artificial Intelligence in Education 17(2):89–120
- Bull S, Kay J (2010) Open learner models. In: Advances in intelligent tutoring systems, Springer, pp 301–322
- Bull S, Ginon B, Boscolo C, Johnson M (2016) Introduction of learning visualisations and metacognitive support in a persuadable open learner model. In: Proc. of Learning Analytics & Knowledge, ACM, pp 30–39
- Carlson R, Genin K, Rau M, Scheines R (2013) Student profiling from tutoring system log data: When do multiple graphical representations matter? In: Proc. of Educational Data Mining
- Carmona C, Millán E, Pérez-de-la Cruz JL, Trella M, Conejo R (2005) Introducing prerequisite relations in a multi-layered bayesian student model. In: User Modeling, Springer, pp 347–356
- Cen H, Koedinger K, Junker B (2006) Learning factors analysis—a general method for cognitive model evaluation and improvement. In: Proc. of Intelligent Tutoring Systems, Springer, pp 164–175
- Cen H, Koedinger KR, Junker B (2007) Is over practice necessary?-improving learning efficiency with the cognitive tutor through educational data mining. Frontiers in Artificial Intelligence and Applications 158:511

- Cen H, Koedinger K, Junker B (2008) Comparing two irt models for conjunctive skills. In: Proc. of Intelligent Tutoring Systems, Springer, pp 796–798
- Chen Y, Wullemin PH, Labat JM (2015) Discovering prerequisite structure of skills through probabilistic association rules mining. In: Proc. of Educational Data Mining, pp 117–124
- Chen Y, Gonzalez-Brenes J, Tian J (2016) Joint discovery of skill prerequisite graphs and student models. In: Proc. of Educational Data Mining
- Chi M, Koedinger K, Gordon G, Jordan P, VanLehn K (2011) Instructional factors analysis. In: Proc. of Educational Data Mining
- Chrysafiadi K, Virvou M (2013) Student modeling approaches: A literature review for the last decade. *Expert Systems with Applications* 40(11):4715–4729
- Conati C, Gertner A, Vanlehn K (2002) Using bayesian networks to manage uncertainty in student modeling. *User modeling and user-adapted interaction* 12(4):371–417
- Corbett AT, Anderson JR (1994) Knowledge tracing: Modeling the acquisition of procedural knowledge. *User modeling and user-adapted interaction* 4(4):253–278
- David YB, Segal A, Gal YK (2016) Sequencing educational content in classrooms using bayesian knowledge tracing. In: Proc. of Learning Analytics & Knowledge, ACM, pp 354–363
- De La Torre J (2009) Dina model and parameter estimation: A didactic. *Journal of Educational and Behavioral Statistics* 34(1):115–130
- Desmarais M (2011) Conditions for effectively deriving a q-matrix from data with non-negative matrix factorization. In: Proc. of Educational Data Mining
- Desmarais M, Lemieux F (2013) Clustering and visualizing study state sequences. In: Proc. of Educational Data Mining
- Desmarais M, Beheshti B, Xu P (2014) The refinement of a q-matrix: Assessing methods to validate tasks to skills mapping. In: Proc. of Educational Data Mining
- Desmarais MC, Baker RS (2012) A review of recent advances in learner and skill modeling in intelligent learning environments. *User Modeling and User-Adapted Interaction* 22(1-2):9–38
- Desmarais MC, Naceur R (2013) A matrix factorization method for mapping items to skills and for enhancing expert-based q-matrices. In: Proc. of Artificial Intelligence in Education, Springer, pp 441–450
- Desmarais MC, Meshkinfam P, Gagnon M (2006) Learned student models with item to item knowledge structures. *User Modeling and User-Adapted Interaction* 16(5):403–434
- Doignon JP, Falmagne JC (2012) Knowledge spaces. Springer Science & Business Media
- Essa A (2016) A possible future for next generation adaptive learning systems. *Smart Learning Environments* 3(1):16
- Falakmasir MH, Pardos ZA, Gordon GJ, Brusilovsky P (2013) A spectral learning approach to knowledge tracing. In: Proc. of Educational Data Min-

ing

- Feng M, Heffernan N, Mani M, Heffernan C (2006) Using mixed-effects modeling to compare different grain-sized skill models. In: Proc. of Educational Data Mining: Papers from the AAAI Workshop
- Gasparetti F, Limongelli C, Sciarrone F (2015) Exploiting wikipedia for discovering prerequisite relationships among learning objects. In: Proc. of Information Technology Based Higher Education and Training, IEEE, pp 1–6
- Goldhammer F (2015) Measuring ability, speed, or both? challenges, psychometric solutions, and what can be gained from experimental control. *Measurement: Interdisciplinary Research and Perspectives* 13(3-4):133–164
- Goldin I, Koedinger K, Alevan V (2013) Hints: You can't have just one. In: Proc. of Educational Data Mining
- Gong Y, Beck J (2011) Items, skills, and transfer models: which really matters for student modeling? In: Proc. of Educational Data Mining
- Gong Y, Beck JE, Heffernan NT (2010) Comparing knowledge tracing and performance factor analysis by using multiple model fitting procedures. In: Proc. of Intelligent Tutoring Systems, Springer, pp 35–44
- Gong Y, Beck JE, Heffernan NT (2011) How to construct more accurate student models: Comparing and optimizing knowledge tracing and performance factor analysis. *International Journal of Artificial Intelligence in Education* 21(1-2):27–46
- Gong Y, Beck JE, Ruiz C (2012) Modeling multiple distributions of student performances to improve predictive accuracy. In: User Modeling, Adaptation, and Personalization, Springer, pp 102–113
- González-Brenes J, Huang Y, Brusilovsky P (2014) 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
- González-Brenes JP, Huang Y (2015) Your model is predictive - but is it useful? theoretical and empirical considerations of a new paradigm for adaptive tutoring evaluation. In: Proc. of Educational Data Mining
- Hämäläinen W, Kumpulainen V, Mozgovoy M (2014) Evaluation of clustering methods for adaptive learning systems. In: Artificial Intelligence Applications in Distance Education, IGI Global
- Hawkins WJ, Heffernan NT, Baker RS (2014) Learning bayesian knowledge tracing parameters with a knowledge heuristic and empirical probabilities. In: Proc. of Intelligent Tutoring Systems, Springer International Publishing, pp 150–155
- Hershkovitz A, de Baker RSJ, Gobert J, Wixon M, Sao Pedro M (2013) Discovery with models a case study on carelessness in computer-based science inquiry. *American Behavioral Scientist* 57(10):1480–1499
- Huang Y, González-Brenes JP, Brusilovsky P (2015a) Challenges of using observational data to determine the importance of example usage. In: Proc. of Artificial Intelligence in Education, Springer, pp 633–637
- Huang Y, González-Brenes JP, Kumar R, Brusilovsky P (2015b) A framework for multifaceted evaluation of student models. In: Proc. of Educational Data Mining

- Junker BW, Sijtsma K (2001) Cognitive assessment models with few assumptions, and connections with nonparametric item response theory. *Applied Psychological Measurement* 25(3):258–272
- Käser T, Busetto AG, Solenthaler B, Baschera GM, Kohn J, Kucian K, von Aster M, Gross M (2013a) Modelling and optimizing mathematics learning in children. *International Journal of Artificial Intelligence in Education* 23(1-4):115–135
- Käser T, Busetto AG, Solenthaler B, Kohn J, von Aster M, Gross M (2013b) Cluster-based prediction of mathematical learning patterns. In: *Proc. of Artificial Intelligence in Education*, Springer, pp 389–399
- Käser T, Klingler S, Schwing AG, Gross M (2014a) Beyond Knowledge Tracing: Modeling Skill Topologies with Bayesian Networks. In: *Proceedings of ITS*, pp 188–198
- Käser T, Koedinger KR, Gross M (2014b) Different parameters - same prediction: An analysis of learning curves. In: *Proc. of Educational Data Mining*, pp 52–59
- Käser T, Klingler S, Gross M (2016) When to stop?: Towards universal instructional policies. In: *Proc. of Learning Analytics & Knowledge*, ACM, pp 289–298
- Khajah M, Wing RM, Lindsey RV, Mozer MC (2014a) Integrating latent-factor and knowledge-tracing models to predict individual differences in learning. In: *Proc. of Educational Data Mining*
- Khajah M, Lindsey RV, Mozer MC (2016) How deep is knowledge tracing? In: *Proc. of Educational Data Mining*
- Khajah MM, Huang Y, González-Brenes JP, Mozer MC, Brusilovsky P (2014b) Integrating knowledge tracing and item response theory: A tale of two frameworks. In: *Workshop on Personalization Approaches in Learning Environments*
- Klingler S, Käser T, Solenthaler B, Gross M (2015) On the Performance Characteristics of Latent-Factor and Knowledge Tracing Models. In: *Proc. of Educational Data Mining*
- Klingler S, Käser T, Solenthaler B, Gross M (2016) Temporally coherent clustering of student data. In: *Proc. of Educational Data Mining*
- Klinkenberg S, Straatemeier M, Van der Maas H (2011) Computer adaptive practice of maths ability using a new item response model for on the fly ability and difficulty estimation. *Computers & Education* 57(2):1813–1824
- Koedinger KR, McLaughlin EA (2016) Closing the loop with quantitative cognitive task analysis. In: *Proc. of Educational Data Mining*
- Koedinger KR, Pavlik Jr PI, Stamper JC, Nixon T, Ritter S (2011) Avoiding problem selection thrashing with conjunctive knowledge tracing. In: *Proc. of Educational Data Mining*, pp 91–100
- Koedinger KR, Corbett AT, Perfetti C (2012) The knowledge-learning-instruction framework: Bridging the science-practice chasm to enhance robust student learning. *Cognitive science* 36(5):757–798
- Koedinger KR, Stamper JC, McLaughlin EA, Nixon T (2013) Using data-driven discovery of better student models to improve student learning. In:

- Proc. of Artificial intelligence in education, Springer, pp 421–430
- Koedinger KR, Yudelson MV, Pavlik PI (2016) Testing theories of transfer using error rate learning curves. *Topics in cognitive science* 8(3):589–609
- Lan AS, Waters AE, Studer C, Baraniuk RG (2014) Sparse factor analysis for learning and content analytics. *Journal of Machine Learning Research* 15(1):1959–2008
- Lee JI, Brunskill E (2012) The impact on individualizing student models on necessary practice opportunities. In: *Proc. of Educational Data Mining*, pp 118–125
- Li N, Cohen W, Koedinger KR, Matsuda N (2011) A machine learning approach for automatic student model discovery. In: *Proc. of Educational Data Mining*
- Lindsey RV, Khajah M, Mozer MC (2014) Automatic discovery of cognitive skills to improve the prediction of student learning. In: *Advances in Neural Information Processing Systems*, pp 1386–1394
- Liu R, Koedinger KR, McLaughlin EA (2014) Interpreting model discovery and testing generalization to a new dataset. In: *Proc. of Educational Data Mining*, pp 107–113
- Liu R, Patel R, Koedinger KR (2016) Modeling common misconceptions in learning process data. In: *Proc. of Learning Analytics & Knowledge, ACM*, pp 369–377
- MacHardy Z, Pardos ZA (2015) Evaluating the relevance of educational videos using bkt and big data. In: *Proc. of Educational Data Mining*
- Mayo M, Mitrovic A (2001) Optimising its behaviour with bayesian networks and decision theory. *Int Journal Artificial Intelligence in Education* 12(2):124–153
- McTavish TS, Larusson JA (2014) Labeling mathematical errors to reveal cognitive states. In: *Open Learning and Teaching in Educational Communities*, Springer, pp 446–451
- Meadows DH (1999) *Leverage points: Places to intervene in a system*. Sustainability Institute Hartland, VT
- Merceron A, Yacef K (2005) Clustering students to help evaluate learning. In: *Technology Enhanced Learning*, Springer, pp 31–42
- Mettler E, Massey CM, Kellman PJ (2011) Improving adaptive learning technology through the use of response times. In: *Proc. of Conference of the Cognitive Science Society*, pp 2532–2537
- Millán E, Loboda T, Pérez-de-la Cruz JL (2010) Bayesian networks for student model engineering. *Computers & Education* 55(4):1663–1683
- Mitrovic A, Martin B (2007) Evaluating the effect of open student models on self-assessment. *International Journal of Artificial Intelligence in Education* 17(2):121–144
- Mitrovic A, Mayo M, Suraweera P, Martin B (2001) Constraint-based tutors: a success story. In: *Proc. of Industrial, Engineering and Other Applications of Applied Intelligent Systems*, Springer, pp 931–940
- Mitrovic A, Koedinger KR, Martin B (2003) A comparative analysis of cognitive tutoring and constraint-based modeling. In: *Proc. of User Modeling*,

- Springer, pp 313–322
- Mitrovic A, Martin B, Suraweera P (2007) Intelligent tutors for all: The constraint-based approach. *IEEE Intelligent Systems* 22:38–45
- Murray RC, Ritter S, Nixon T, Schwiebert R, Hausmann RG, Towle B, Fancsali SE, Vuong A (2013) Revealing the learning in learning curves. In: *Proc. of Artificial Intelligence in Education*, Springer, pp 473–482
- Nakic J, Granic A, Glavinic V (2015) Anatomy of student models in adaptive learning systems: A systematic literature review of individual differences from 2001 to 2013. *Journal of Educational Computing Research* 51(4):459–489
- Nižnan J, Pelánek R, Papoušek J (2015a) Exploring the role of small differences in predictive accuracy using simulated data. In: *AIED Workshop on Simulated Learners*
- Nižnan J, Pelánek R, Řihák J (2015b) Student models for prior knowledge estimation. In: *Proc. of Educational Data Mining*, pp 109–116
- Nixon T, Fancsali S, Ritter S (2013) The complex dynamics of aggregate learning curves. In: *Proc. of Educational Data Mining*
- Nižnan J, Pelánek R, Řihák J (2014) Mapping problems to skills combining expert opinion and student data. In: *Proc. of Mathematical and Engineering Methods in Computer Science*, Springer, LNCS, vol 8934, pp 113–124
- Ohlsson S (1994) Constraint-based student modeling. In: *Student modelling: the key to individualized knowledge-based instruction*, Springer, pp 167–189
- Papoušek J, Pelánek R, Stanislav V (2014) Adaptive practice of facts in domains with varied prior knowledge. In: *Proc. of Educational Data Mining*, pp 6–13
- Papoušek J, Stanislav V, Pelánek R (2016) Evaluation of an adaptive practice system for learning geography facts. In: *Proc. of Learning Analytics & Knowledge*, ACM
- Pardos ZA, Heffernan NT (2009) Determining the significance of item order in randomized problem sets. In: *Proc. of Educational Data Mining*
- Pardos ZA, Heffernan NT (2010a) Modeling individualization in a bayesian networks implementation of knowledge tracing. In: *Proc. of User Modeling, Adaptation, and Personalization*, Springer, pp 255–266
- Pardos ZA, Heffernan NT (2010b) Navigating the parameter space of bayesian knowledge tracing models: Visualizations of the convergence of the expectation maximization algorithm. In: *Proc. of Educational Data Mining*, pp 161–170
- Pardos ZA, Heffernan NT (2011) Kt-idem: Introducing item difficulty to the knowledge tracing model. *User Modeling, Adaption and Personalization* pp 243–254
- Pardos ZA, Xu Y (2016) Improving efficacy attribution in a self-directed learning environment using prior knowledge individualization. In: *Proc. of Learning Analytics & Knowledge*, ACM, pp 435–439
- Pardos ZA, Heffernan NT, Anderson B, Heffernan CL, Schools WP (2010) Using fine-grained skill models to fit student performance with bayesian networks. *Handbook of educational data mining* 417

- Pardos ZA, Gowda SM, Baker RS, Heffernan NT (2012a) The sum is greater than the parts: ensembling models of student knowledge in educational software. *ACM SIGKDD explorations newsletter* 13(2):37–44
- Pardos ZA, Trivedi S, Heffernan NT, Sárközy GN (2012b) Clustered knowledge tracing. In: *Proc. of Intelligent Tutoring Systems*, Springer, pp 405–410
- Pardos ZA, Wang QY, Trivedi S (2012c) The real world significance of performance prediction. *Proc of Educational Data Mining*
- Pavlik PI, Anderson JR (2005) Practice and forgetting effects on vocabulary memory: An activation-based model of the spacing effect. *Cognitive Science* 29(4):559–586
- Pavlik PI, Anderson JR (2008) Using a model to compute the optimal schedule of practice. *Journal of Experimental Psychology: Applied* 14(2):101
- Pavlik PI, Cen H, Koedinger KR (2009) Performance factors analysis—a new alternative to knowledge tracing. In: *Proc. of Artificial Intelligence in Education*, IOS Press, pp 531–538
- Pavlik PI, Brawner K, Olney A, Mitrovic A (2013) A review of student models used in intelligent tutoring systems. In: Sottolare RA, Graesser A, Hu X, Holden H (eds) *Design Recommendations for Intelligent Tutoring Systems*, vol 1, US Army Research Laboratory, pp 39–68
- Pavlik Jr P, Bolster T, Wu SM, Koedinger K, Macwhinney B (2008) Using optimally selected drill practice to train basic facts. In: *Proc. of Intelligent Tutoring Systems*, Springer, pp 593–602
- Pelánek R (2014) Application of time decay functions and elo system in student modeling. In: *Proc. of Educational Data Mining*, pp 21–27
- Pelánek R (2015) Metrics for evaluation of student models. *Journal of Educational Data Mining* 7(2)
- Pelánek R (2015) Modeling students’ memory for application in adaptive educational systems. In: *Proc. of Educational Data Mining*, pp 480–483
- Pelánek R (2016) Applications of the elo rating system in adaptive educational systems. *Computers & Education* 98:169–179
- Pelánek R, Jarušek P (2015) Student modeling based on problem solving times. *International Journal of Artificial Intelligence in Education* pp 1–27
- Pelánek R, Řihák J (2016) Properties and applications of wrong answers in online educational systems. In: *Proc. of Educational Data Mining*
- Pelánek R, Řihák J, Papoušek J (2016) Impact of data collection on interpretation and evaluation of student model. In: *Proc. of Learning Analytics & Knowledge*, ACM, pp 40–47
- Piech C, Bassen J, Huang J, Ganguli S, Sahami M, Guibas LJ, Sohl-Dickstein J (2015) Deep knowledge tracing. In: *Advances in Neural Information Processing Systems*, pp 505–513
- Qiu Y, Qi Y, Lu H, Pardos ZA, Heffernan NT (2011) Does time matter? modeling the effect of time with bayesian knowledge tracing. *EDM* pp 139–148
- Reddy S, Labutov I, Banerjee S, Joachims T (2016) Unbounded human learning: Optimal scheduling for spaced repetition. In: *Proc. of Knowledge Discovery and Data Mining*, ACM

- Ritter S, Yudelson M, Fancsali SE, Berman SR (2016) How mastery learning works at scale. In: Proc. of Learning@Scale, ACM, pp 71–79
- Rollinson J, Brunskill E (2015) From predictive models to instructional policies. In: Proc. of Educational Data Mining
- Rubin DC, Hinton S, Wenzel A (1999) The precise time course of retention. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 25(5):1161
- van de Sande B (2013) Properties of the bayesian knowledge tracing model. *Journal of Educational Data Mining* 5(2):1
- Sao Pedro M, Baker RS, Gobert JD (2013a) Incorporating scaffolding and tutor context into bayesian knowledge tracing to predict inquiry skill acquisition. In: Proc. of Educational Data Mining, pp 185–192
- Sao Pedro MA, Baker RS, Gobert JD (2013b) What different kinds of stratification can reveal about the generalizability of data-mined skill assessment models. In: Proc. of Learning Analytics and Knowledge, ACM, pp 190–194
- Scheines R, Silver E, Goldin I (2014) Discovering prerequisite relationships among knowledge components. In: Proc. of Educational Data Mining, pp 355–356
- Sculley D, Holt G, Golovin D, Davydov E, Phillips T, Ebner D, Chaudhary V, Young M, Crespo JF, Dennison D (2015) Hidden technical debt in machine learning systems. In: *Advances in Neural Information Processing Systems*, pp 2494–2502
- Sense F, Behrens F, Meijer RR, Rijn H (2016) An individual’s rate of forgetting is stable over time but differs across materials. *Topics in cognitive science* 8(1):305–321
- Smith AC, Frank LM, Wirth S, Yanike M, Hu D, Kubota Y, Graybiel AM, Suzuki WA, Brown EN (2004) Dynamic analysis of learning in behavioral experiments. *The journal of neuroscience* 24(2):447–461
- Stamper JC, Koedinger KR, McLaughlin EA (2013) A comparison of model selection metrics in datashop. In: Proc. of Educational Data Mining, pp 284–287
- Stone M (1977) An asymptotic equivalence of choice of model by cross-validation and akaike’s criterion. *Journal of the Royal Statistical Society Series B (Methodological)* pp 44–47
- Straatemeier M (2014) Math garden: A new educational and scientific instrument. PhD thesis, Universiteit van Amsterdam, Faculty of Social and Behavioural Sciences
- Streeter M (2015) Mixture modeling of individual learning curves. In: Proc. of Educational Data Mining
- Tang S, McBride E, Gogel H, Pardos ZA (2015) Item ordering effects with qualitative explanations using online adaptive tutoring data. In: Proc. of Learning@Scale, ACM, pp 313–316
- Tatsuoka KK (1983) Rule space: An approach for dealing with misconceptions based on item response theory. *Journal of Educational Measurement* 20(4):345–354

- Thai-Nghe N, Drumond L, Horváth T, Krohn-Grimberghe A, Nanopoulos A, Schmidt-Thieme L (2011) Factorization techniques for predicting student performance. *Educational Recommender Systems and Technologies: Practices and Challenges* pp 129–153
- Toscher A, Jahrer M (2010) Collaborative filtering applied to educational data mining. *KDD cup*
- Trivedi S, Pardos ZA, Heffernan NT (2011) Clustering students to generate an ensemble to improve standard test score predictions. In: *Proc. of Artificial Intelligence in Education*, Springer, pp 377–384
- Valdés Aguirre B, Ramírez Uresti JA, Boulay Bd (2016) An analysis of student model portability. *International Journal of Artificial Intelligence in Education* 26(3):932–974
- Van Der Linden W (2009) Conceptual issues in response-time modeling. *Journal of Educational Measurement* 46(3):247–272
- Van Inwegen EG, Adjei SA, Wang Y, Heffernan NT (2015) Using partial credit and response history to model user knowledge. In: *Proc. of Educational Data Mining*
- Vanlehn K (2006) The behavior of tutoring systems. *International journal of artificial intelligence in education* 16(3):227–265
- Vanlehn K, Lynch C, Schulze K, Shapiro JA, Shelby R, Taylor L, Treacy D, Weinstein A, Wintersgill M (2005) The andes physics tutoring system: Lessons learned. *International Journal of Artificial Intelligence in Education* 15(3):147–204
- Řihák J (2015) Use of time information in models behind adaptive system for building fluency in mathematics. In: *Proc. of Educational Data Mining, Doctoral Consortium*
- Řihák J, Pelánek R (2016) Choosing a student model for a real world application. In: *Building ITS Bridges Across Frontiers (ITS Workshop)*
- Wang X, Berger JO, Burdick DS, et al. (2013) Bayesian analysis of dynamic item response models in educational testing. *The Annals of Applied Statistics* 7(1):126–153
- Wang Y, Beck J (2013) Class vs. student in a bayesian network student model. In: *Proc. of Artificial Intelligence in Education*, Springer, pp 151–160
- Wang Y, Heffernan N (2013) Extending knowledge tracing to allow partial credit: Using continuous versus binary nodes. In: *Proc. of Artificial Intelligence in Education*, Springer, pp 181–188
- Wang Y, Heffernan NT, Heffernan C (2015) Towards better affect detectors: effect of missing skills, class features and common wrong answers. In: *Proc. of Learning Analytics And Knowledge*, ACM, pp 31–35
- Wauters K, Desmet P, Van Den Noortgate W (2012) Item difficulty estimation: An auspicious collaboration between data and judgment. *Computers & Education* 58(4):1183–1193
- White KG (2001) Forgetting functions. *Animal Learning & Behavior* 29(3):193–207
- Xu Y, Mostow J (2012) Comparison of methods to trace multiple subskills: Is lr-dbn best?. In: *Proc. of Educational Data Mining*

- Yudelson MV, Koedinger KR (2013) Estimating the benefits of student model improvements on a substantive scale. In: Educational Data Mining Workshops Proceedings
- Yudelson MV, Koedinger KR, Gordon GJ (2013) Individualized bayesian knowledge tracing models. In: Proc. of Artificial Intelligence in Education, Springer, pp 171–180

Radek Pelánek received his Ph.D. degree in Computer Science from Masaryk University for his work on formal verification. Since 2010 his research interests focus on areas of educational data mining and learning analytics. Currently he is the leader of the Adaptive Learning group at Masaryk University and is interested in both theoretical research in user modeling and practical development of adaptive educational systems.