Chapter (X): Knowledge Inference Models Used in Adaptive Learning

Maria Ofelia Z. San Pedro, ***ACT, Inc.***
Ryan S. Baker, ***University of Pennsylvania***

Abstract

This chapter provides an overview of adaptive learning and examines the student model component used in adaptive learning systems. Established and more recent approaches to student modeling that infer student knowledge (i.e. what students know at any given moment during the learning experience) are discussed, as student knowledge is the most common learner characteristic widely assessed in large-scale adaptive systems. This chapter concludes with a discussion of the limitations of the current generation of adaptive learning systems, and areas of potential for future progress.

Keywords: Adaptive learning, Student Knowledge, Bayesian Knowledge Tracing, Performance Factor Analysis, Deep Knowledge Tracing, Elo Rating

In recent years, digital environments for learning and assessment have come to provide rich opportunities to capture and utilize data on the learning experiences of students, increasing the potential to fully personalize learning in these environments. When online learning systems use data to tailor to student characteristics and performance, they become adaptive learning environments. Although existing systems generally fall short of this idea (see review in Baker, 2016), an ideal adaptive learning environment would adapt to assessments of a wide range of student attributes and constructs influential to the learning experience – student knowledge, academic emotions, behavioral engagement and motivation. Successful adaptive learning systems can utilize detailed learner activity data in real-time to make inferences which are in turn used to automatically provide the learner with individualized support. Such support can come in the form of *adaptive sequencing* in the student's learning experience by automatically adjusting the sequence of the next content or skills to master (VanLehn (2006)'s "outer-loop"), including adjusting difficulty of the items given, and also in *adaptive content* in response to a student's unique response, such as corrective feedback, context-sensitive hints, supportive messages, metacognitive advice (VanLehn's "inner-loop")..

Examples of adaptive learning systems that have successfully emerged over the years include intelligent tutoring systems (ITSs), adaptive platforms developed by educational publishers, adaptive educational games, adaptive educational hypermedia systems, and adaptive online courseware. The combination of maturing technology and government educational initiatives has led to a growth in the number of adaptive learning systems used by schools. Increasingly, ITSs designed to teach math skills are being used at large-scale (hundreds of thousands of users each year, in schools and districts across the US), including Cognitive Tutors from Carnegie Learning, shown to lead to better performance on standardized exams (Pane, Griffin, McCaffrey, & Karam, 2014), and subsequent math courses (Koedinger, Corbett, & Ritter, 2000); and ALEKS from McGraw Hill Education, shown to increase class attendance and math performance when implemented in after-school programs (Craig et al., 2011). Some systems have been designed to provide a platform for instructor users to develop and import their content and pedagogical practices (e.g. Smart Sparrow, CogBooks, ASSISTments), while others provide instructional content created by the provider (e.g. Adapt Courseware, LearnSmart, Reasoning Mind).

The development of adaptive learning systems has in many cases leveraged methods from learning analytics (LA) and educational data mining (EDM) (Baker & Siemens, 2014), communities that employ machine learning and statistical techniques in the processing and modeling of fine-grained data. An alternate paradigm, Computational Psychometrics (CP), aims to integrate data-driven computer science methods (machine learning and data mining, in particular) in theoretical psychometric frameworks to support accurate and valid measurements of latent abilities in real time (Von Davier, 2017). Computational psychometrics attempts to address data dependencies and multidimensionality that are not easily handled by traditional psychometric models (e.g., IRT), using psychometric approaches towards handling the hierarchy in educational data. Both EDM/LA and CP have the goal of measuring learners' proficiencies, skills, and other attributes. A comprehensive discussion of the differences between EDM/LA and CP is, however, outside the scope of this chapter.

In this chapter, we provide an overview of adaptive learning and describe its essential components. In particular, this chapter will focus on several student modeling approaches that evaluate *what students know* (i.e. student knowledge) at any given moment during the learning experience, as this is the most common learner characteristic widely assessed in large-scale adaptive systems (Pelanek, 2017a). We then look into how these approaches are themselves validated, and discuss examples of how adaptive systems tailor learner experiences based on those assessments. We conclude with a discussion of the limitations of the current generation of adaptive learning systems, and areas of potential for future progress.

What is Adaptive Learning?

Adaptive learning in educational settings is a method for personalizing learning experiences for students. Adaptive learning uses technology (typically online systems) and the data the system collects to evaluate students' interactions and performance towards identifying their needs and tailoring instruction, remediation, and other aspects of the learning experience to aid students to learn the material and succeed. The goal is to provide the best possible learning content and resources for the students' needs at a particular point in time. While most extant systems used at scale adapt to students' knowledge and limitations in understanding, research systems have adapted to other aspects as well, such as affect and motivation (Arroyo et al., 2014; D'Mello et al., 2010; Rebolledo-Medez, du Boulay, & Luckin, 2005). Adaptive learning leverages findings and methods from cognitive science, educational psychology and artificial intelligence, in order to achieve these goals. Modern adaptive learning systems can identify what a student knows and does not know, their skills, and strategies, and can pin down the source of the misconceptions. They can in many other cases detect a student's learning goals, affective state, level of engagement, and their motivation to learn.

It is important to note that adaptive learning is different from adaptive assessment or testing. As mentioned earlier, adaptive testing estimates the student's proficiency to select the next test item to present to the student. The goal of adaptive testing is to maximize the amount of information about proficiency gained from the test while minimizing the amount of time required to complete the test. In contrast, the goal of adaptive learning is to optimize the student's learning experience by both assessing his or her proficiency or behavior and providing the necessary individualized support towards mastery and achievement. This is similar to the framework of Reinforcement Learning (RL), which models sequential decisions where an agent interacting with an environment learns a policy (a mapping of states to action) that maximizes a reward (Sutton and Barto, 1998). In the context of learning and instruction, these states can be students' cognitive states, the actions are instructional activities (e.g., problems, worked examples, video content, etc.) that can modify the student's cognitive state, and the reward can be a measure of the quality of the action (e.g., reward for a student learning a skill).

One core aspect of most adaptive learning systems is the explicit use of a student model (Sottilare, Holden, Graesser, & Hu, 2013; Wenger, 1987; Woolf, 2010) that consists of measures of the student's knowledge, goals, engagement, and other pertinent attributes that enable the system to make informed decisions in adapting to individual students. (Note that this use of the term "student model" is different from the use seen in the Evidence-Centered Design framework, i.e. Mislevy & Riconscente, 2006, and precedes it by a considerable amount of time). Student models can contain assessments of each of these constructs, and update them in real time, so that they can be used in adaptation or provided to the instructor. It is worthwhile to note that adaptive learning systems can be designed to use explicit student models first when deciding on a specific adaptation to a student (e.g., model-based RL) or only use data to learn an adaptation without any student models (e.g., model-free RL), as we discuss later in this chapter.

Framework for Adaptive Learning Systems

At their core, online or digital learning systems are made adaptive by being built with three major components: a domain model, a student model, and a pedagogical or instructional model (Figure 1). This is broadly the same set of components that has been used to describe intelligent tutoring systems (ITSs) (Sottilare, Holden, Graesser, & Hu, 2013; Wenger, 1987; Woolf, 2010), and similar to the components used to describe adaptive educational hypermedia (AEH) (Brusilovsky, Karagiannidis, & Sampson, 2004; Karampiperis & Sampson, 2005). The *domain model* refers to the content domain to be taught by the system, containing the set of topics and the corresponding learning objectives or outcomes, skills, knowledge, and strategies needed to learn them (Nkambou, Mizoguchi, & Bourdeau, 2010; Sottilare, Holden, Graesser, & Hu, 2013; Woolf, 2010). It is a representation of the ideal expert knowledge which can be in the form of content standards, prerequisite knowledge, or common misconceptions that students normally exhibit for a skill or content. The *student model* assesses the student users' cognitive state, as well as other states (such as affective, behavioral and motivational states) based on the student's interaction with the learning system. The information from the student model is what adaptive learning systems consider in selecting the content and instruction to present to the student user. The *pedagogical or instructional model* selects the specific content, tutoring strategies and feedback for an individual student at a specific time (generally designed based on theories in pedagogy and instruction; in RL, it is referred algorithmically as policy or the instructional policy). It takes information from the domain and student models as inputs and decides the next activity the system will provide to the student user based on that information (Sottilare, Holden, Graesser, & Hu, 2013; VanLehn, 2006). While this component allows for automated adaptation in terms of deciding what to do next, certain mixed-initiative systems allow students to initiate the actions, ask questions or request for help (Aleven, McLaren, Roll & Koedinger, 2006).

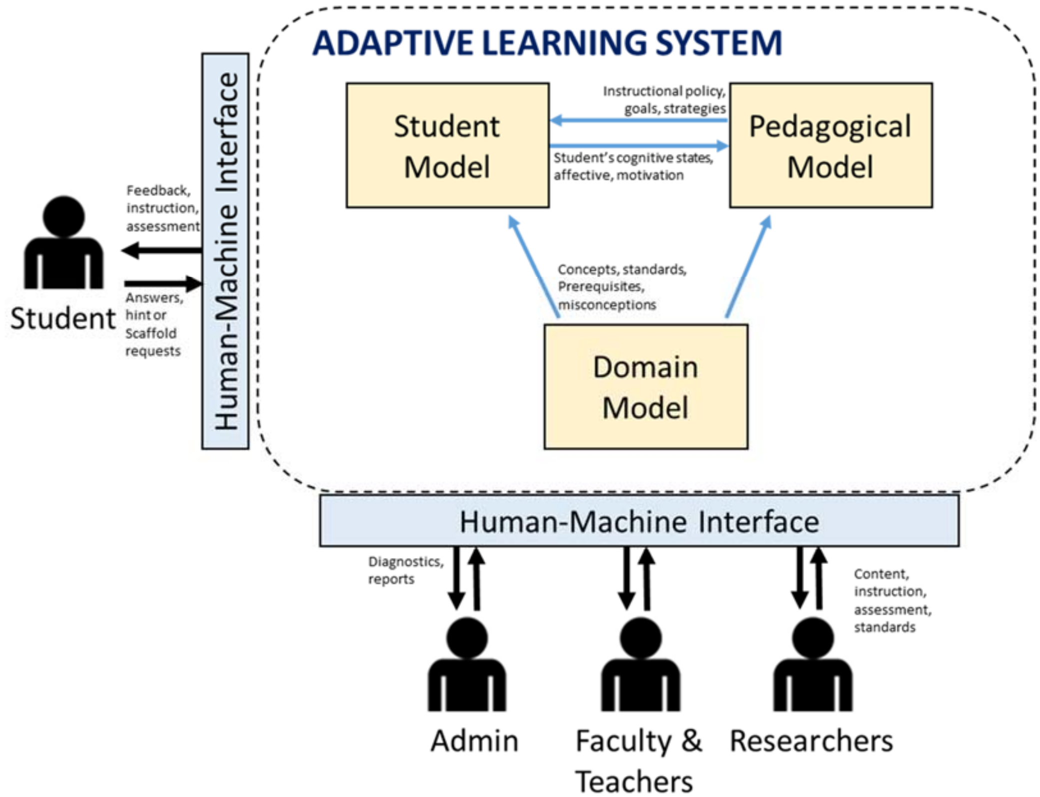

Figure 1. Adaptive Learning Systems Framework.

Adapting to Student Knowledge

Developing a good student model is usually seen as the starting point in the development of adaptive learning systems. There has been considerable research to determine which learner attributes can form the basis of meaningful and effective intervention, while at the same time being tractable to capture within a real-world system (Ahn & Brusilovsky, 2009; Biswas et al., 2005; Canfield, 2001; Graesser et al., 2007; Hu et al., 2012; Jeremic, Jovanovic, & Gasevic, 2009; Koedinger et al., 2006; Murray & Arroyo, 2002; Razzaq et al., 2005; VanLehn, 2006). While there are several approaches to adaptability, most of current implemented adaptive learning systems are characterized by inferring student knowledge, and then adapting in an immediate fashion to that inference, for example by providing a hint to a struggling student or by providing mastery learning – giving the student additional practice on a difficult skill until he or she demonstrates mastery. Learning systems that adapt based on machine-learning based detection of student affect or meta-cognition, while successful in limited studies (Azevedo et al., 2012; D'Mello et al., 2009; DeFalco, et al., 2018), have not yet scaled the way that knowledge-based adaptivity has. Systems using mastery learning represent student knowledge as skills, facts, or concepts, and track student progress from one problem or activity to the next, building a skill mastery profile for the students in order to eventually make decisions about when the student should move on to the next topic.

One core example of an adaptive learning system that uses knowledge estimates within mastery learning is the Cognitive Tutor (Koedinger & Corbett, 2006), mentioned above. The system makes use of a probabilistic user model that assesses student knowledge in real-time. With each student action (e.g., answering correctly or incorrectly, requesting for help) linked to a skill (or sometimes referred to as a knowledge component or KC) needed to solve a problem, the system evaluates these actions and aggregates these evaluations into evidence about which skills the student has learned or not learned (Corbett & Anderson, 1995), and displays this skills progress to students and teachers. The system then provides the student with practice on the current skill until the student demonstrates mastery of that skill.

Andes is an adaptive learning system that teaches Newtonian physics using coached problem solving (VanLehn, 1996) where the system and the student collaborate in solving a problem. Its student model uses a Bayesian Network probabilistic framework that assesses a student's plans in solving the problem, predicts the student's goals and actions, and provides a long-term evaluation of a student's domain knowledge. The tutor updates its student model based on the problems the student has worked on and the student's fine-grained interactions with the system during problem solving, assessing the probability and uncertainty of the student's knowledge state (Conati, Gertner, VanLehn, & Druzdzel, 1997).

Another example of an adaptive learning system that evaluates student knowledge in its student model is the SQL-Tutor, an intelligent tutoring system used in courses on database theory and the SQL database query language (Mitrovic, 1998). It selects a problem based on a constraint-based student model (Ohlsson, 1994) that represents correct knowledge in solving a problem as a set of constraints – aspects a correct solution must have. If a student's solution violates a constraint, the system intervenes through feedback attached to that constraint. This feedback informs the student what is incorrect, why it was incorrect and points out the corresponding domain principle attached to that constraint.

By modeling what a student knows, and updating a student's skill mastery profile in real-time, the system can then organize instruction around what the student knows, as well as provide feedback on the student's current state of knowledge. There have been several modeling approaches to infer what a student knows at a given time. In this section, we review one widely-used knowledge inference model, and then discuss other recent approaches to inferring student knowledge.

Bayesian Knowledge Tracing

To model and quantify student learning, many adaptive learning systems employ an approach known as Bayesian Knowledge Tracing (BKT) (Corbett & Anderson, 1995). BKT aims to infer the latent construct of the student's knowledge state for a skill (i.e. does a student know a certain skill at a given time?) from a student's pattern of correct and incorrect answers to problems or problem steps that require that skill up until that point. The model posits that at any given moment to demonstrate a skill, the knowledge state of a student is either mastered or not mastered, with a certain probability that is calculated given the student's observed performance of correct and incorrect responses (Figure 2). It updates its knowledge inference that the student has mastered a skill or not on each opportunity of demonstrating that skill (e.g., through an item or problem step). Some extensions to Bayesian Knowledge Tracing have also added probabilistic correctness on items as well as binary correctness (right, wrong) (e.g. Sao Pedro et al., 2013).

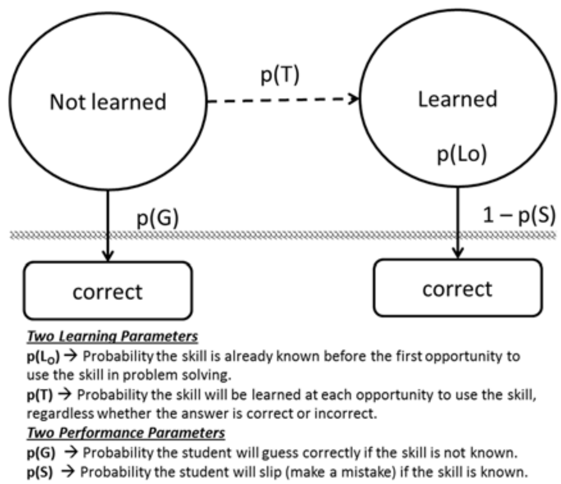

Figure 2. Corbett and Anderson's model of Bayesian Knowledge Tracing.

In its original formulation, only the first attempt at an opportunity to practice or demonstrate a skill (sometimes referred to as a practice opportunity) is considered, and each item or problem step corresponds to only a single skill. A set of four parameters is fit for each skill, and these four parameters remain invariant across the entire context of using the learning system: two learning parameters, $P(L_0)$ or initial knowledge (initial probability of knowing the skill) and P(T) or probability of learning the skill (at each opportunity to make use of the skill), together with two performance parameters P(G) or guess (probability that the student will give a correct answer despite not knowing a skill) and P(S) or slip (probability that the student will give an incorrect answer despite knowing the skill). Both the expectation maximization algorithm and grid search have been used to estimate the parameters from training data. The assumption is made that students do not forgot a skill once they have learned it. Another assumption is that all items, within the same skill, have the same difficulty, and that no contextual factors (beyond what skill it is) impact student performance or learning. Reye (2004) shows how BKT equations could be derived from a Hidden Markov Model or Dynamic Bayesian Network. Several algorithms have been used to estimate the parameters from training data, with approximately equal performance; several packages are available on the web to fit BKT parameters.

Using Bayesian analysis below, the system continually updates the estimate that a student knows a skill, whenever a student gives a first response to an item or problem step (correct or incorrect, bug, help request; help requests are treated as evidence that the student does not know the skill). This is done by first re-calculating the probability that the student knew the skill before making that first attempt (equations 1 and 2), and then updating the probability based on the possibility that the student learned the skill at the current item or problem step (equation 3).

$$P(L_{n-1} \mid Correct_n) = \frac{P(L_{n-1}) * (1 - P(S))}{P(L_{n-1}) * (1 - P(S)) + (1 - P(L_{n-1})) * P(G)} \qquad (1)$$

$$P(L_{n-1} \mid Incorrect_n) = \frac{P(L_{n-1}) * P(S)}{P(L_{n-1}) * P(S) + (1 - P(L_{n-1})) * (1 - P(G))} \qquad (2)$$

$$P(L_n \mid Action_n) = P(L_{n-1} \mid Action_n) + ((1 - P(L_{n-1} \mid Action_n)) * P(T)) \qquad (3)$$

Two common parameter fitting methods for standard BKT[1] uses Expectation-Maximization (EM)[2] and brute force grid search[3].

One of the key assumptions of BKT is that its parameters vary by skill or knowledge components but remain constant for all other factors. In recent years, significant research have been conducted to remove or modify this assumption – either by relaxing one of its assumptions or considering additional information. For example, researchers have contextualized the parameter estimates of guess and slip using number of attempts and time spent (Baker, Corbett, & Aleven, 2008), contextualized the four original parameters with the use of help (Beck et al., 2008), individualized the initial probability of mastery (prior per-student) (Pardos & Heffernan, 2010), added item-level parameters difficulty in BKT (Pardos & Heffernan, 2011; Khajah et al., 2014; Gonzalez-Brenes et al., 2014), and have accounted for student differences in terms of initial mastery probabilities and skill learning probabilities (Yudelson et al., 2013).

Performance Factor Analysis

Another relatively prominent model for student knowledge inference in ITSs is Performance Factor Analysis (PFA) (Pavlik, Cen, & Koedinger, 2009). In PFA, the student's skill estimate is represented by estimating the probability that a student will get the item related to that skill correct. PFA employs a logistic regression model to predict student performance for a new item in terms of the number of success and failures (correctness or incorrectness) attained for the skills involved in that item so far. Unlike BKT, PFA assumes each item may involve multiple latent traits or skills. Each skill in the PFA model has the following parameters:

---

[1] A scalable code written in C/C++ that fits Hidden Markov Models (BKT specifically) at scale can be found at https://github.com/myudelson/hmm-scalable

[2] EM used in modeling BKT with Bayes Net Toolbox for Student Modeling (BNT-SM). Matlab code can be found at http://www.cs.cmu.edu/~listen/BNT-SM/

[3] Brute force grid search program code written in Java can be found at http://www.upenn.edu/learninganalytics/ryanbaker/edmtools.html

β the difficulty of the skill (can also represent item difficulty in some model variants)
γ the effect of prior successes tracked s for a skill (or KC), or success learning rate
ρ the effect of prior failures tracked f for a skill (or KC), or failure learning rate

Together with tracking the prior successes (s) and prior failures (f) for the KC for the student on each relevant skill, the probability that the learner will get the next item correct ($P(m)$)is computed (with Expectation Maximization used to fit the parameters):

$$m(i, j \in KCs, s, f) = \sum_{j \in KCs}(\beta_j + \gamma_j s_{i,j} + \rho_j f_{i,j}) \quad (4)$$

$$P(m) = \frac{1}{1 + e^{-m}} \quad (5)$$

In published comparisons, PFA has achieved predictive performance roughly comparable to BKT (see review in Pardos et al., 2011).

Deep Knowledge Tracing

Deep knowledge tracing (DKT), introduced by Piech et al. (2015), is based on deep learning or recurrent neural networks (RNN) (Rumelhart et al., 1985) to represent the latent knowledge state. Instead of producing a separate model for each skill, DKT models all skills together. The input to the RNNs consists of compressed representations of a student action (one-hot-encodings), resulting in a vector of predictions on the probability of getting items or exercises in the dataset correct. By using the sequence of student performance across skills, DKT provides predictions of a student's performance on future items within the system. Aside from the input and output layers, the RNN model also has a hidden layer with recurrent connections that serves to retain relevant aspects of the input that are useful for predicting future performance (Khajah, et al., 2016). As a student progresses, DKT utilizes information from previous items when useful to make inferences regarding future performance in every skill. RNNs have a high dimensional, continuous, representation of latent state giving them the ability to use this rich information as input in making prediction at a much later point in time. DKT has thus far been trained using stochastic gradient descent on mini batches, optimizing for negative log likelihood.

Khajah and colleagues (2016) found that modern extensions to BKT (inclusion of forgetting, latent student abilities and skill induction) achieved performance equivalent to DKT, though both approaches had higher predictive accuracy than Classical BKT. However, more recent extensions to DKT have proven better than BKT and PFA both at predicting future performance within the learning system (Zhang et al., 2017; Yeung & Yeung, 2018) and at predicting future performance on paper post-tests (Scruggs et al., 2020).

Elo Rating System

An emerging paradigm for predicting student knowledge in adapting learning systems is the Elo Rating System (Pelanek, 2016). Traditional Item Response Theory (IRT) assesses a student's knowledge of a particular topic by computing the probability that a student will get an item correct based on a fixed ability parameter for the student and difficulty and (sometimes) discriminability parameters of the item. It assumes there is only one skill being measured per set of items, though other psychometrics models relax this assumption (e.g., CDM, Rupp, Templin & Henson, 2010; multidimensional IRT, Embretson & Reise, 2000). IRT also assumes that no learning occurs between items, a problematic assumption in online adaptive learning, where a lack of learning would imply that the learning system was fundamentally failing in its goals. The Elo Rating System (Pelanek, 2016), a variant of the 1-PL IRT model, relaxes this assumption and can be used in a running system: the Elo Rating System (Pelanek, 2016) which is. The Elo Rating System continually estimates item difficulty and student ability, updating both every time a student encounters an item. It was originally devised for chess rating but has been used in student modeling by evaluating a student's item response as a "match" between the student and the item.

Applying the Elo rating system in the context of adaptive learning systems, Pelanek (2016) denotes the skill of a student *s* as θ_S, item (*i*) difficulty as d_i and the student response as $correct_{si} \in \{0, 1\}$. A logistic function using the skill and item difficulty gives the probability of a correct answer:

$$P(correct_{si} = 1) = \frac{1}{1 + e^{-(\theta_S - d_i)}} \qquad (6)$$

$$\theta_S \coloneqq \theta_S + K \cdot (correct_{si} - P(correct_{si} = 1) \quad (7)*$$
$$d_i \coloneqq d_i + K \cdot (P(correct_{si} = 1) - correct_{si} \quad (8)*$$

* *where K is a parameter for how strongly the model should consider new information.*

The Elo rating system has been shown to provide good prediction accuracy (Nižnan, Pelánek, & Rihák, 2015), and better performance than standard PFA and BKT when extended in combination with PFA (Papousek, Pelánek, & Stanislav, 2014). The Elo rating system provides both simplicity and flexibility in its student modeling approach which makes it easy to implement in educational systems (Pelanek, 2016; Von Davier et al., 2019; Yudelson, Rosen, Polyak & de la Torre, 2019).

Model Assessment

Models such as BKT, PFA, DKT, and Elo are often assessed and validated in multiple ways. The most common way that these types of models are assessed is by building the model on one group of students, applying it to a new, unseen group of students, and testing the model's performance at predicting future performance for these students. Typically, performance on the n^{th} opportunity to demonstrate a skill is predicted from the model's state at the $(n-1)^{th}$ opportunity to demonstrate that skill. Most commonly, student-level cross-validation is used so that every student's data can be used in both training and test set. By doing so, it becomes possible to determine whether a model is generalizable. The resultant predictions are evaluated using multiple metrics – there is an ongoing and continuing debate in the field as to which of several metrics is best, including AUC ROC, RMSE, log likelihood, and others (Pardos & Yudelson, 2013; Pelanek, 2015). There is also debate as to whether evaluation should be conducted at the grain-size of all data points together (i.e. differentiating prediction across students and skills) or individual within-student or within-skill (i.e. differentiating the model's ability to determine when a student learns or when a specific skill is known) (Pelanek, 2017b). Note that a model's ability to differentiate whether a student has learned – by being able to predict when a student will demonstrate successful or unsuccessful performance -- is somewhat different than common psychometric notions of reliability of assessment in that the knowledge states of students are changing as they are being assessed. If the student is learning – and in an adaptive learning system we hope they are – we should not hope that our assessments of that student will agree over time.

Although the most common form of assessment of this type of knowledge model is immediate future performance, many papers have also looked at whether the resultant models could predict performance on future tests of the same knowledge, administered through other modalities, such as paper post-tests (see, for instance, Corbett & Anderson, 1995; Sao Pedro et al., 2013). This practice can help to establish that the model is capturing a more general indicator of student knowledge. Models are often also studied to determine whether all items belong within the skill, through examining whether better fit is achieved if an item is excluded (Corbett & Anderson, 1995; Stamper & Koedinger, 2011). One other key differentiator between models is the degree to which they infer a latent construct. Whereas BKT and Elo infer a latent construct (i.e., student knowledge, ability), which can be used in several ways, DKT and PFA are limited to predicting performance on specific items. This limitation means that, despite DKT's better performance than BKT's original articulation, it cannot be straightforwardly used for many of the purposes that BKT has been used for, such as predicting performance outside the learning system and use in models of engagement and affect (e.g. Pardos et al., 2014).

Adapting to Student Attributes

Once learner attributes (in this paper, student knowledge) are modeled by an adaptive learning system, the pedagogical model is designed to select problems or customize instructional content for the learner. Adaptive systems take into account the information in the student models and behave differently in terms of pedagogical mechanisms for different students or groups of students, aimed at providing individualized support for better learning.

An example of adaptation that is more complex than simple mastery learning is seen in Assessment and Learning in Knowledge Spaces (ALEKS), (Canfield, 2001; Hu et al., 2012) a web-based assessment and learning system that attempts to identify when a student is struggling because the student does not know a prerequisite skill, so that it can switch to supporting the student in learning the prerequisite skill. As the student progresses in learning topics, ALEKS updates its student model and uses that information to make a principled selection of what skills it asks the student to work on.

Several aspects of the learning experience can be adapted, including the content's complexity or difficulty (Murray & Arroyo, 2002; VanLehn, 2006), scaffolding and support (Razzaq et al., 2005; Koedinger et al., 2006), the visual interface (Ahn & Brusilovsky, 2009; Jeremic, Jovanovic, & Gasevic, 2009), and feedback prompts such as hints, prompts or communications from pedagogical agents (Biswas et al., 2005; Razzaq et al., 2005; Graesser et al., 2007). While many of these systems have hand-coded intervention logic, some experimental systems go beyond this to use algorithms which attempt to automatically discover which types of interventions work for which students, under which conditions (e.g. Chi, VanLehn, Litman, & Jordan, 2011). Through this process, the system "learns" more about the student in real-time, and what works for this student, as the student responds to the instructional or learning material presented to them.

Finally, some systems provide support through natural language dialogues. The best known example is AutoTutor (Graesser, Jackson, & McDaniel, 2007). It employs semantic analyses to compare a student's input text to the text of the ideal answer to a problem, as well as sample text representing other potential answers, including both good answers and misconceptions. By recognizing a range of potential answers (and being able to do textual transformations between canonical answers and alternate ways of saying the same thing), AutoTutor can respond in a flexible and sensitive fashion. It adaptively responds to the student answers, whether answers are incomplete or representing misconceptions, by engaging in a variety of tutorial strategies, including asking for more information, giving positive, neutral or negative messages, prompts for missing words, giving hints, answering questions from students, and summarizing answers (Graesser, Jackson, & McDaniel, 2007).

Conclusion

In this chapter, we provided an overview of student modeling used in adaptive learning, focusing on modern approaches to inferring student knowledge – the most common student attribute being adapted to in current large-scale adaptive learning systems. We have discussed a small number of established adaptive educational systems, as well as established and more recent knowledge inference models. Although not discussed in detail here, most of the systems presented in this chapter have been shown to lead to significantly better student outcomes than traditional curricular approaches (Graesser, Jackson, & McDaniel, 2007; Hu et al., 2012; Koedinger, Anderson, Hadley, & Mark, 1997; Leelawong & Biswas, 2008; Razzaq, et al., 2005), although relatively few systems have achieved a high rating in research review databases such as the What Works Clearinghouse or Evidence for ESSA. Especially in the United States, adaptive learning systems have achieved scale, with systems such as the Cognitive Tutor or ALEKS being used by hundreds of thousands of students.

One interesting limitation to the current generation of adaptive learning systems, discussed in detail in (Baker, 2016), is that with a few exceptions (such as Affective AutoTutor, for instance), most existing systems adapt to a single dimension of the learner, in a single way. Individual systems become very effective at using mastery learning, or at providing feedback when students violate constraints, and perhaps giving messages as to why the student is wrong – but comprehensive support for problem selection, misconceptions, affect, motivation, disengaged behavior, and self-regulated learning is rare. Baker (2016) suggests that this is because of the difficulty and complexity of getting any one type of intervention correct; for adaptive learning systems to reach their full potential, they will have to surmount this challenge and consider and adapt to the student in a richer, more multi-dimensional fashion.

References

Ahn, J. W., & Brusilovsky, P. (2009). Adaptive visualization of search results: Bringing user models to visual analytics. *Information Visualization*, 8(3), 167-179.

Aleven, V., McLaren, B., Roll, I. & Koedinger, K. R. (2006). Toward meta-cognitive tutoring: A model of help seeking with a cognitive tutor. *International Journal of Artificial Intelligence in Education*, 16, 101-128.

Arroyo, I., Woolf, B. P., Burelson, W., Muldner, K., Rai, D., & Tai, M. (2014). A multimedia adaptive tutoring system for mathematics that addresses cognition, metacognition and affect. *International Journal of Artificial Intelligence in* Education, 24(4), 387-426.

Azevedo, R., Landis, R. S., Feyzi-Behnagh, R., Duffy, M., Trevors, G., Harley, J. M., ... & Yeasin, M. (2012). The effectiveness of pedagogical agents' prompting and feedback in facilitating co-adapted learning with MetaTutor. In *International conference on intelligent tutoring systems* (pp. 212-221). Springer, Berlin, Heidelberg.

Baker, R.S. (2016) Stupid Tutoring Systems, Intelligent Humans. *International Journal of Artificial Intelligence and Education, 26* (2), 600-614.

Baker, R. S. J. d., Corbett, A. T., & Aleven, V. (2008). More accurate student modeling through contextual estimation of slip and guess probabilities in bayesian knowledge tracing. In E. Aimeur & B. Woolf (Eds.), In *Proceedings of the 9th International Conference on Intelligent Tutoring Systems* (pp. 406–415).

Baker, R., Siemens, G. (2014) Educational data mining and learning analytics. In Sawyer, K. (Ed.) *Cambridge Handbook of the Learning Sciences: 2nd Edition,* pp. 253-274.

Beck, J., Chang, K. M., Mostow, J., & Corbett, A. (2008). Does help help? Introducing the Bayesian Evaluation and Assessment methodology. In *Intelligent Tutoring Systems* (pp. 383-394). Springer Berlin/Heidelberg.

Biswas, G., Leelawong, K., Schwartz, D., Vye, N., & The Teachable Agents Group at Vanderbilt. (2005). Learning by teaching: A new agent paradigm for educational software. *Applied Artificial Intelligence*, 19(3-4), 363-392.

Brusilovsky, P., Karagiannidis, C., & Sampson, D. (2004). Layered evaluation of adaptive learning systems. *International Journal of Continuing Engineering Education and Life Long Learning*, *14*(4-5), 402-421.

Canfield, W. (2001). ALEKS: A Web-based intelligent tutoring system. *Mathematics and Computer Education*, 35(2), 152.

Chi, M., VanLehn, K., Litman, D., & Jordan, P. (2011). An evaluation of pedagogical tutorial tactics for a natural language tutoring system: A reinforcement learning approach. *International Journal of Artificial Intelligence in Education*, 21(1), 83-113.

Conati, C., Gertner, A. S., VanLehn, K., & Druzdzel, M. J. (1997, January). On-line student modeling for coached problem solving using Bayesian networks. In *User Modeling* (pp. 231-242). Springer Vienna.

Corbett, A.T. & Anderson, J.R. (1995) Knowledge Tracing: Modeling the Acquisition of Procedural Knowledge. *User Modeling and User-Adapted Interaction*, 4 (4), 253-278.

Craig, S. D., Anderson, C., Bargagloitti, A., Graesser, A. C., Okwumabua, T., Sterbinsky, A., & Hu, X. (2011, January). Learning with ALEKS: The Impact of Students' Attendance in a Mathematics After-School Program. In *Artificial Intelligence in Education* (pp. 435-437).

D'Mello, S., Craig, S., Fike, K., & Graesser, A. (2009). Responding to learners' cognitive-affective states with supportive and shakeup dialogues. In *International Conference on Human-Computer Interaction* (pp. 595-604). Springer, Berlin, Heidelberg.

D'Mello, S., Lehman, B., Sullins, J., Daigle, R., Combs, R., Vogt, K., ... & Graesser, A. (2010). A time for emoting: When affect-sensitivity is and isn't effective at promoting deep learning. *In Intelligent tutoring systems* (pp. 245-254).

DeFalco, J.A., Rowe, J.P., Paquette, L., Georgoulas-Sherry, V., Brawner, K., Mott, B.W., Baker, R.S., Lester, J.C. (2018) Detecting and Addressing Frustration in a Serious Game for

Military Training. *International Journal of Artificial Intelligence and Education, 28* (2), 152-193.
Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. Mahwah, NJ: Erlbaum.
González-Brenes, J., Huang, Y., & Brusilovsky, P. (2014). General features in knowledge tracing to model multiple subskills, temporal item response theory, and expert knowledge. In *The 7th International Conference on Educational Data Mining* (pp. 84-91).
Graesser, A.C., Jackson, G.T., & McDaniel, B. (2007). AutoTutor holds conversations with learners that are responsive to their cognitive and emotional states. *Educational Technology*, 47, 19-22.
Hu, X., Craig, S. D., Bargagliotti, A. E., Graesser, A. C., Okwumabua, T., Anderson, C., ... & Sterbinsky, A. (2012). The Effects of a Traditional and Technology-based After-school Setting on 6th Grade Student's Mathematics Skills. *Journal of Computers in Mathematics and Science Teaching*, 31(1), 17-38.
Jeremic, Z., Jovanovic, J., & Gaševic, D. (2009). Evaluating an intelligent tutoring system for design patterns: The DEPTHS experience. *Educational Technology & Society*, 12(2), 111-130.
Karampiperis, P., & Sampson, D. (2005). Adaptive Learning Resources Sequencing in Educational Hypermedia Systems. *Educational Technology & Society*, 8 (4), 128-147.
Khajah, M. M., Huang, Y., González-Brenes, J. P., Mozer, M. C., & Brusilovsky, P. (2014). Integrating knowledge tracing and item response theory: A tale of two frameworks. In *Proceedings of Workshop on Personalization Approaches in Learning Environments (PALE 2014) at the 22th International Conference on User Modeling, Adaptation, and Personalization* (pp. 7-12).
Khajah, M., Lindsey, R. V., & Mozer, M. C. (2016). How deep is knowledge tracing?. *arXiv preprint arXiv:1604.02416*.
Koedinger, K. R., Anderson, J. R., Hadley, W. H., & Mark, M. A. (1997). Intelligent tutoring goes to school in the big city. *International Journal of Artificial Intelligence in Education*, 8, 30-43
Koedinger, K. R. & Corbett, A. T. (2006). Cognitive Tutors: Technology bringing learning science to the classroom. In K. Sawyer (Ed.) *The Cambridge Handbook of the Learning Sciences*, (pp. 61-78).
Koedinger, K. R., Corbett, A. T., & Ritter, S. (2000). *Carnegie Learning's Cognitive Tutor: Summary Research Results*.
Leelawong, K., & Biswas, G. (2008). Designing learning by teaching agents: The Betty's Brain system. *International Journal of Artificial Intelligence in Education*, 18(3), 181-208.
Mislevy, R. J., & Riconscente, M. M. (2006). Evidence-centered assessment design. *Handbook of test development*, 61-90.
Mitrovic, A. (1998). A knowledge-based teaching system for SQL. *In Proceedings of ED-MEDIA (*Vol. 98, pp. 1027-1032).
Murray, T., Arroyo, I. (2002) Toward Measuring and Maintaining the Zone of Proximal Development in Adaptive Instructional Systems. *Proceedings of the International Conference on Intelligent Tutoring Systems,* 749-758.
Nižnan, J., Pelánek, R., & Rihák, J. (2015). Student Models for Prior Knowledge Estimation. *International Educational Data Mining Society*.

Nkambou, R., Mizoguchi, R., & Bourdeau, J. (Eds.). (2010). *Advances in intelligent tutoring systems* (Vol. 308). Springer Science & Business Media.

Ohlsson, S.: 1994, Constraint-based Student Modeling. *Student Modeling: the Key to Individualized Knowledge-based Instruction*, pp. 167-189.

Pane, J. F., Griffin, B. A., McCaffrey, D. F., & Karam, R. (2014). Effectiveness of cognitive tutor algebra I at scale. *Educational Evaluation and Policy Analysis*, *36*(2), 127-144.

Papousek, J., Pelánek, R., & Stanislav, V. (2014, July). Adaptive practice of facts in domains with varied prior knowledge. In *Educational Data Mining 2014.*

Pardos, Z. A., & Yudelson, M. (2013, July). Towards Moment of Learning Accuracy. In *AIED Workshops.*

Pardos, Z.A., Baker, R.S.J.d., Gowda, S.M., Heffernan, N.T. (2011) The Sum is Greater than the Parts: Ensembling Models of Student Knowledge in Educational Software. *SIGKDD Explorations, 13* (2), 37-44.

Pardos, Z. A., Baker, R. S., San Pedro, M., Gowda, S. M., & Gowda, S. M. (2014). Affective States and State Tests: Investigating How Affect and Engagement during the School Year Predict End-of-Year Learning Outcomes. *Journal of Learning Analytics*, 1(1), 107-128.

Pardos, Z. A., & Heffernan, N. T. (2010). Modeling individualization in a bayesian networks implementation of knowledge tracing. In User Modeling, Adaptation, and Personalization (pp. 255-266). Springer Berlin Heidelberg.

Pardos, Z., & Heffernan, N. (2011). KT-IDEM: introducing item difficulty to the knowledge tracing model. User Modeling, Adaption and Personalization, 243-254.

Pavlik Jr, P. I., Cen, H., & Koedinger, K. R. (2009). Performance Factors Analysis--A New Alternative to Knowledge Tracing. *Online Submission*.

Pelánek, R. (2015). Metrics for evaluation of student models. *Journal of Educational Data Mining,* 7(2), 1-19.

Pelánek, R. (2016). Applications of the Elo rating system in adaptive educational systems. *Computers & Education*, 98, 169-179.

Pelánek, R. (2017a). Bayesian knowledge tracing, logistic models, and beyond: an overview of learner modeling techniques. *User Modeling and User-Adapted Interaction*, *27*(3-5),313-350.

Pelánek, R. (2017b). Measuring Predictive Performance of User Models: The Details Matter. In *Adjunct Publication of the 25th Conference on User Modeling, Adaptation and Personalization (UMAP '17)*. ACM, New York, NY, USA, 197-201.

Piech, C., Bassen, J., Huang, J., Ganguli, S., Sahami, M., Guibas, L. J., & Sohl-Dickstein, J. (2015). Deep knowledge tracing. In *Advances in Neural Information Processing Systems* (pp. 505-513).

Razzaq, L., Feng, M., Nuzzo-Jones, G., Heffernan, N. T., Koedinger, K. R., Junker, B., ... & Rasmussen, K. P. (2005). The Assistment project: Blending assessment and assisting. In *Proceedings of the 12th Annual Conference on Artificial Intelligence in Education* (pp. 555-562).

Rebolledo-Mendez, G., du Boulay, B., & Luckin, R. (2005, May). Be bold and take a challenge": Could motivational strategies improve help-seeking. In Proceedings of the *2005 conference on Artificial Intelligence in Education: Supporting Learning through Intelligent and Socially Informed Technology* (pp. 459-466).

Reye, J. (2004). Student modelling based on belief networks. *International Journal of Artificial Intelligence in Education*, *14*(1), 63-96.

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1985). *Learning internal representations by error propagation* (No. ICS-8506). California Univ San Diego La Jolla Inst for Cognitive Science.

Rupp, A. A., Templin, J., & Henson, R. A. (2010). *Diagnostic measurement: Theory, methods, and applications*. Guilford Press.

Sao Pedro, M.A., Baker, R.S.J.d., Gobert, J., Montalvo, O. Nakama, A. (2013) Leveraging Machine-Learned Detectors of Systematic Inquiry Behavior to Estimate and Predict Transfer of Inquiry Skill. *User Modeling and User-Adapted Interaction*, 23 (1), 1-39.

Scruggs, R., Baker, R.S., McLaren, B.M. (2020) Extending Deep Knowledge Tracing: Inferring Interpretable Knowledge and Predicting Post System Performance. *Proceedings of the 28th International Conference on Computers in Education*.

Sottilare, R., Holden, H., Graesser, A. C., & Hu, X. (2013). *Design Recommendations for Adaptive Intelligent Tutoring Systems: Learner Modeling (Vol. 1)*. US Army Research Laboratory.

Stamper, J., & Koedinger, K. (2011). Human-machine student model discovery and improvement using DataShop. In *Artificial intelligence in education* (pp. 353-360). Springer Berlin/Heidelberg.

Sutton, R. S. and Barto, A. G. (1998). *Reinforcement learning: An introduction.* MIT Press.

VanLehn, K. (1996). Conceptual and meta learning during coached problem solving. In *Intelligent Tutoring Systems* (pp. 29-47). Springer Berlin Heidelberg.

VanLehn, K. (2006) The Behavior of Tutoring Systems*. International Journal of Artificial Intelligence and Education*, 16 (3), 227-265.

Von Davier, A. A. (2017). Computational Psychometrics in Support of Collaborative Educational Assessments. *Journal of Educational Measurement*, *54*(1), 3-11.

Von Davier, A. A., Deonovic, B. E., Yudelson, M., Polyak, S., & Woo, A. (2019). Computational psychometrics approach to holistic learning and assessment systems. In *Frontiers in Education* (Vol. 4, p. 69).

Wenger, E. (1987). *Artificial Intelligence and Tutoring Systems. Computational and Cognitive Approaches to the Communication of Knowledge*. Loas Altos, CA: Morgan Kaufmann

Woolf, B. P. (2010). *Building intelligent interactive tutors: Student-centered strategies for revolutionizing e-learning*. Morgan Kaufmann.

Yeung, C.-K., & Yeung, D.-Y. (2018). Addressing two problems in deep knowledge tracing via prediction-consistent regularization. In *Proceedings of the Fifth Annual ACM Conference on Learning at Scale*, 1–10.

Yudelson, M. V., Koedinger, K. R., & Gordon, G. J. (2013, July). Individualized bayesian knowledge tracing models. In *International Conference on Artificial Intelligence in Education* (pp. 171-180). Springer, Berlin, Heidelberg.

Yudelson, M., Rosen, Y., Polyak, S., de la Torre (2019). Leveraging skill hierarchy for multi-level modeling with Elo rating System. In *Proceedings of the Sixth ACM Conference on Learning @ Scale*. IL: Chicago.

Zhang, J., Shi, X., King, I., and Yeung, D.-Y. (2017). Dynamic key-value memory networks for knowledge tracing. In *Proceedings of the 26th International Conference on World Wide Web*, 765–774.