Dai, M., Hung, J., Du, X., Tang H. & Li, H. (2021).Knowledge tracing: A review of available techniques. *Journal of Educational Technology Development and Exchange*, *14*(2), 1-20.

Knowledge Tracing: A Review of Available Techniques

Miao Dai
Central China Normal University, China
daimiao@mails.ccnu.edu.cn

Jui-Long Hung
Central China Normal University, China, Boise State University, USA
andy.edtech@gmail.com

Xu Du
Central China Normal University, China
duxu@mail.ccnu.edu.cn

Hengtao Tang
University of South Carolina, USA
htang@mailbox.sc.edu

Hao Li
Central China Normal University, China
lihao205@mail.ccnu.edu.cn

Abstract: *As a student modeling technique, knowledge tracing is widely used by various intelligent tutoring systems to infer and trace the individual's knowledge state during the learning process. In recent years, various models were proposed to get accurate and easy-to-interpret results. To make sense of the wide Knowledge tracing (KT) modeling landscape, this paper conducts a systematic review to provide a detailed and nuanced discussion of relevant KT techniques from the perspective of assumptions, data, and algorithms. The results show that most existing KT models consider only a fragment of the assumptions that relate to the knowledge components within items and student's cognitive process. Almost all types of KT models take "quize data" as input, although it is insufficient to reflect a clear picture of students' learning process. Dynamic Bayesian network, logistic regression and deep learning are the main algorithms used by various knowledge tracing models. Some open issues are identified based on the analytics of the reviewed works and discussed potential future research directions.*

Keywords: knowledge tracing, assumptions, data, algorithm

1. Introduction

An intelligent tutoring system (ITS, hereafter) is a computer system that aims to provide immediate and customized instruction or feedback to learners (Freedman et al., 2000). Online education is becoming popular in today's educational systems; integrating ITS into online or blended learning has attracted significant research efforts (Hilles & Naser, 2017). The major goal of ITS aims to provide in-time and personalized feedback, how to monitor and evaluate student's learning progress and performance is the most important component for triggering corresponding actions. The domain of learning progress tracking and evaluation is so-called "Knowledge Tracing" (KT, hereafter)—a family of algorithms to model each learner's mastery of the knowledge being tutored (Corbett & Anderson, 1994).

The learning of a student is a process of acquiring new knowledge through interaction with the external learning environment on the basis of his/her prior knowledge (Appleton & Beasley, 1994). The algorithms of KT are to infer and trace the individual's "knowledge state" during the learning process, that is, the exact set of concepts mastered by the individual. Ideally, the knowledge state results generated should be in-time, accurate, and easy-to-interpret to enable follow up teaching and learning decisions (Kasurinen & Nikula, 2009; V. Swamy et al., 2018) and educational recommendations(Han et al., 2016). To achieve the above goals, many models have been proposed by researchers, such as Bayesian knowledge tracing (BKT, hereafter) (Corbett & Anderson, 1994), performance factors analysis (PFA, hereafter) (Pavlik Jr et al., 2009), and deep knowledge tracing (DKT, hereafter) (Piech et al., 2015).

It is quite difficult to be certain about a student's knowledge state within a KT model because of the uncertainty of his/her learning process (Vlahavas & Spyropoulos, 2002). Students' mastery of the target domain knowledge is influenced not only by their general prior knowledge but also by the learning context (Self, 1990). Ideally, all of such contextual information should be represented within the KT model, so that ITS can provide suitable interactive assistance to students. To make sense of the wide KT modeling landscape, we argue that there is a need to propose a scientific paradigm to simplify and aggregate available techniques. Such a scientific paradigm will provide basic guidelines for researchers to understand the roles of current techniques and identify areas that require further clarification in future research.

The first component of the scientific paradigm for KT modeling that we consider relates to the theoretical assumptions. Assumptions are implemented by various KT models to conclude data. However, some key assumptions of commonly used KT modeling formalisms may not be valid (Yudelson et al., 2008) and it is often unclear whether underlying assumptions of any commonly used formalism will necessarily hold true (Gong et al., 2011). Checking model assumptions could optimize model performance and increase model reliability.

The second component of the scientific paradigm for KT modeling relates to the source of data. Based on the available observational data about a student's interaction with an ITS, KT models estimate the current state of a student's knowledge and provide a prediction of future performance. Students' correctness attempts on certain knowledge components (KCs, hereafter)[1], which will be referred to as "quiz data" in this paper, are the basic data source for KT modeling. The performance on quizzes[2] is used to model student knowledge growth. To enhance the model's accuracy and stability, many rich and informative data such as students' longitudinal electroencephalography (EEG, hereafter) signals (Xu et al., 2014), multi-behavior features (Lap Pong & Haiqin, 2017; Sun et al., 2019), and temporal information (Zhu et al., 2018) were proposed for incorporation and yielded better results than using quiz data alone. To give a better bird's eye view of current work, there is a need to assess what data about learners and learning environments can be collected and used for KT modeling and what results can be concluded from these research efforts.

The third and last component of the scientific paradigm for KT modeling relates to algorithms. Although model assumptions have the potential to prove general results, these results depend critically on the form of algorithms used. Many efforts have been made focusing on introducing new algorithms and proving they were superior to previous ones (e.g. dynamic Bayesian network, logistic regression, recurrent neural network). However, is there any generalizable model(s) that can be applied in assorted circumstances? If not, it would be beneficial to summarize individual algorithms' strengths and weaknesses.

All the above research gaps need to be addressed by analyzing various KT models systematically. Therefore, the purpose of this study aims to conduct a literature review, mainly focusing on KT models from the perspective of KT modeling techniques including assumptions, data, and algorithms. Our research efforts aim to answer the following questions:

RQ1. What are the characteristics of publications in KT?

RQ2. What are the general assumptions of a KT model? How might existing assumptions influence a KT model?

RQ3. Based on the literature, what data could be adopted by a KT model? What results can be concluded from these research efforts?

RQ4. What are the strengths and weaknesses of major knowledge tracing algorithms? What new research needs are generated by these new approaches?

2. Background

2.1 Knowledge Tracing

KT is one of student modeling techniques which has attracted intensive research efforts. The task of KT can be formulated as a supervised learning problem: given a student's past interactions $X_T=(x_1,x_2,\ldots,x_t)$ up to time t on a particular learning task, the performance of a student is predicted in the next interaction x_{t+1}. An interaction $x_t=(q_t,a_t)$ is defined as a tuple containing the KC id q_t of a question that a student attempts at time step t, and the label a_t is a binary variable that represents whether the student answers correctly or not.

1. Knowledge component (KC) is a generalization of everyday terms like concept, principle, fact, or skill, and cognitive science term like schema, production rule, misconception or facet.

2. This paper will interchangeably refer to quizzes as questions, items, exercises or problems.

KT usually seeks to predict the probability that the student will answer the question correctly during the next time-step, i.e., $p(a_{t+1}=1|q_{t+1},X_T)$.

In ITSs, KT models have two common usages. To predict students' performance in the next practice opportunity is the most frequently used one. For example, BKT is used in ACT Programming Tutor to predict students' knowledge mastery during problem practicing (Corbett & Anderson, 1994). Mongkhonvanit et al. (2019) utilized a DKT framework to predict a student's next item response with over 88% accuracy in MOOC. The other usage of KT models in ITSs is to obtain explainable parameter estimates (Gong et al., 2011). Being explainable indicates the parameters produced by the KT model have practical meanings (i.e. pinpoint intuitively which KCs a student is good at or unfamiliar with), which can help researchers know more about scientific facts. For example, Schodde et al. (2017) adapted proactive instruction to students in a game-like tutoring interaction by interpreting parameters estimated from BKT. Jin et al. (2019) proposed a recommendation algorithm to match suitable exercises to students adaptively based on understanding BKT parameters.

Overall, various KT models are used in assorted circumstances for users to make informed, valid decisions. The modeling processes and potential pros and cons of the KT model are scattered in various studies. Therefore, it is necessary to conduct a systematic review to provide a detailed and nuanced discussion of relevant KT techniques.

2.2 Major Models for Knowledge Tracing

Corbett and Anderson (1994) first used a 2-node Dynamic Bayesian Network to model the knowledge state of each KC separately for each student in ITS, and proposed the model called BKT. As the dominant method of modeling student knowledge, BKT has the characteristics of simplicity, accurate prediction, and ease of interpretation. However, it cannot capture learning where multiple skills are needed to perform a single action (Gong et al., 2011). To handle multiple KCs for the same item, Pavlik Jr et al. (2009) presented a new alternative KT model called PFA. PFA uses a logistic function to predict the probability of correctness and is somewhat superior to BKT. Considering the significant correlation of factors (e.g. KCs) to diverse learning states, several works shed light on the possibility of knowledge state computing using deep learning algorithms such as DKT using Long Short-Term Memory (LSTM) (Piech et al., 2015), Dynamic Key-Value Memory Networks (DKVMN) using Memory-Augmented Neural Networks (MANN) (Jiani et al., 2016), graph-based knowledge tracing (GKT) using Graph Neural Network (GNN) (Nakagawa et al., 2019) and so on.

In all, multiple variants, extensions, and alternatives to KT models have been developed based on the strengthening of the theoretical framework and the adoption of new algorithms. In this paper, we will take an in-depth look at those KT models, what data is required of them, why we need them with regards to different learning pedagogies, and what algorithms are identified to distinguish between different KT models.

3. Methodology

3.1 Literature Search Strategy

The purpose of this study is to investigate techniques that have been used in KT research and generate summative findings to answer our research questions. Related articles were extracted from the following databases: IEEE Xplore Digital Library, ACM Digital Library, and Web of Science. “Knowledge tracing/ leaner modeling/student modeling” was used as the research term to extract studies that used it in the title or as a key term, and the search time is 2019. The range of the article selection is from 1/1994 to 8/2019 and 139 references were initially retrieved from those 3 databases.

3.2 Inclusion/exclusion Criteria

The main purpose of this study is to discuss KT modeling techniques; the following criteria were set to identify the articles to be included in the analysis. (1) Articles that defined KT as the primary research goal; (2) articles that provide detailed information related to KT models/algorithms/techniques; (3) articles that focused on proposing new methods to generate a more sensitive and accurate estimation of student knowledge state, rather than an implementation; (4) articles that focused on solving issues existing in KT modeling techniques. Reviews, commentaries, and case studies were excluded from the data analysis of this study. Finally, 48 papers were included and then coded by the classification scheme in the next section.

3.3 The Classification Schemes

To give a better bird’s-eye view of current work, we introduced our classification schemes. It is based on the inclusion/exclusion criteria for classifying KT algorithms, and includes the extensively studied families of probabilistic graph model, logistic regression model, and deep learning model. The individual dimensions are shown in Table 1.

Table 1
The classification schemes.

Dimensions	Explanation
Probabilistic graph model	Models representing students’ knowledge state using probability distributions and can be derived from a reasonable dynamic Bayesian Network (DBN), including BKT, multiple variants within BKT per se, and extensions to BKT.
Logistic regression model	Models that are based on a logistic regression function.
Deep learning model	Models that employed deep learning algorithms.

4. Results

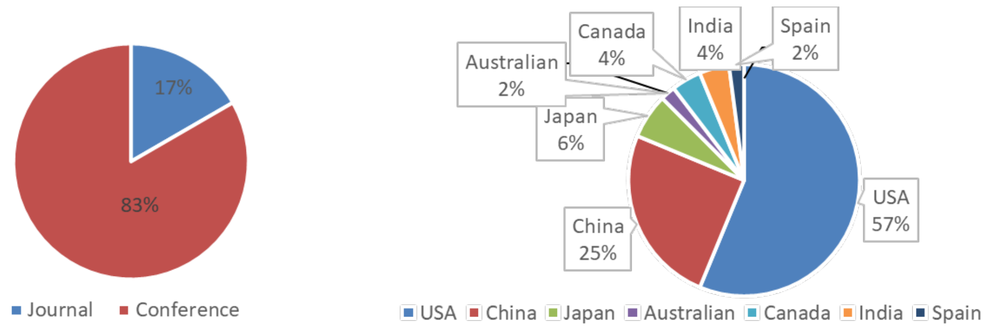
In this section, results are extracted from the selected studies starting with bibliometrics, then moving to the assumptions, data, and algorithms of various KT models.

4.1 Bibliometrics

Here, we conduct a basic statistical analysis of the origin of the selected studies and trends of publication numbers.

4.1.1 Origin of the selected studies

Based on the inclusion/exclusion criteria, 48 papers were selected after careful examination according to the classification scheme. Figure 1(a) illustrates the distribution of publications that were selected. Around 83% (40/48) of studies from 1995 to 2019 are conference papers, while only 17% (8/48) are journal articles. This may result from the databases we chose. Figure 1(b) shows that the first authors of all publications came from seven countries and more than half of scholars are from the USA (57% (27/48)), followed by China (25% (12/48)) and Japan (6% (3/48)).



(a) (b)

Figure 1
Illustration of the selected studies: (a) The distribution of publications; (b) The distribution of first authors' nationalities.

4.1.2 Trends of publication numbers.

Figure2 shows the number of publications across years; modeling approaches are coded according to the classification scheme we proposed. The x-axis represents the published year of papers and the y-axis represents the number of papers. Most studies belong to the family of a probabilistic graph model, while the number of papers on the logistic model is least. It is worthwhile to notice that studies on deep learning models show an increasing trend since 2015.

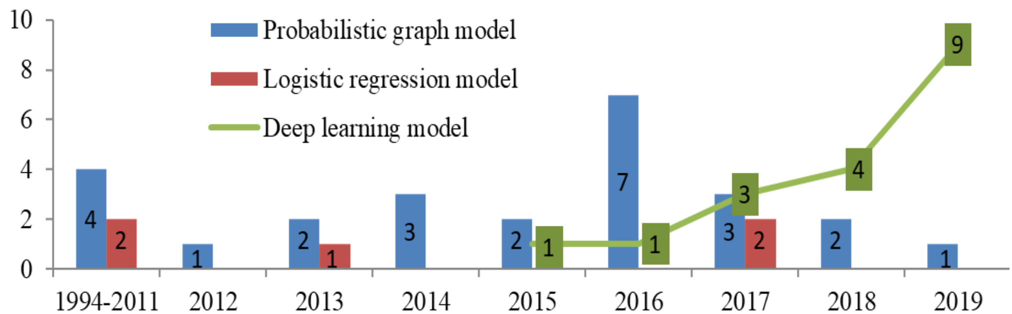


Figure 2
Modeling approaches of the selected studies.

4.2 Assumptions of Knowledge Tracing Models

In this section, we discuss the assumptions behind various KT models. When using the KT model to make actual decisions, it is important to be aware of the differences between the model and reality, and their implications.

In the traditional classroom, paper-and-pencil pretests or posttests in various forms are used to assess the knowledge level of students. The test results only report a general score, but candidates with the same score often have different knowledge states and different cognitive structures. Therefore, instructors are increasingly dissatisfied with getting only individual macro-level evaluations, and they would like tests to provide specific and personalized diagnostic information, especially to reflect the cognitive structure of the students (e.g. what knowledge the candidate has mastered), and put forward corresponding suggestions (e.g. which aspects of the student still need to be strengthened); this is what we call cognitive diagnosis.

Applying the KT model to student personalized counseling and cognitive diagnosis is an important step forward. Assumptions from cognitive diagnoses are used to simplify the modeling process and highlight the interplay between model inputs and outputs. Based on the literature, assumptions of various KT models are extracted from each article selected and shown in Table 2. The significance in Table 2 means that KT models are based on a set of core assumptions from items/KCs, students' personality and their learning process.

Table 2

Assumptions extracted for each article selected. The numbers in brackets represent the total of studies.

Dimensions	Explanation
Assumptions about students	Different students come with different characteristics (16).
Assumptions about items/ KCs	One item covers one/multiple KC(s) that present in a course (48).
	Assuming conditional independence of all KCs (34).
	There are rich structures and correlations among different KCs of a learning domain (14).
	Different items have different difficulty or helpfulness levels for students (4).
Assumptions about learning	If a student knows the KC, she/he would correctly perform and if a student correctly performs, then she/he has acquired that KC (8).
	As students interact with items, they learn the KCs that are presented in them(48).
	Students' knowledge in these KCs will increase with frequent repetitions but will gradually lose under the influence of time (13).

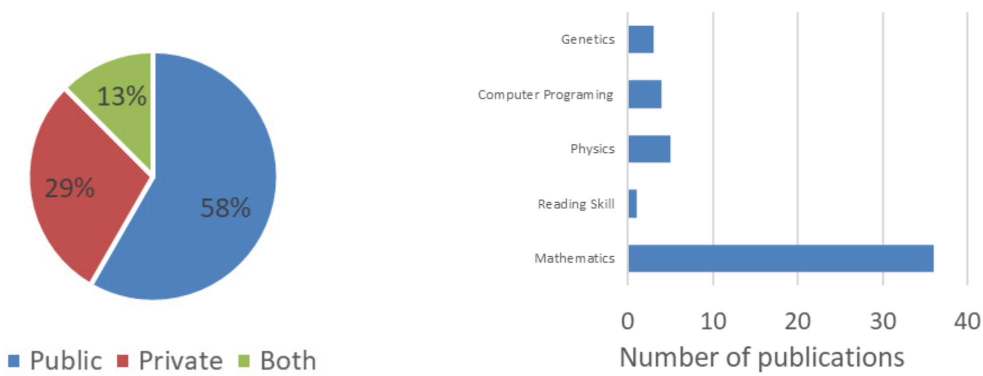
	Once a student is in the known state, she/he doesn't transit to an unknown state (25).
Assumptions about learning	Characterizing students' knowledge state as to whether they have mastered a certain KC or not (33).
	Characterizing students' performance as multiple state observable variables or continuous partial credit (15).

4.3 Data for Knowledge Tracing Modeling

This section outlines what kind of data one may need to build a KT model, and who provides such data.

4.3.1 Data sources.

The data sources are coded in Fig 3(a). From the 48 studies reviewed, around 58% (28/48) used public datasets only. In particular, the public dataset used is mainly from the KDD Cup Educational Data Mining Challenge[3], the ASSISTments platform[4], and Khan Academy[5]. These datasets consist of answers to student's historical questions and most of the questions are classified in domain knowledge, which are available free on the web. As shown in Fig 3(b), there are 5 kinds of domain-specific subjects among all the datasets, and mathematics was the most.



(a) (b)

Figure 3

Illustration of the selected studies. (a) Dataset sources; (b) Number of publications per domain.

4.3.2 Types of data.

KT models include a range of model types including the probabilistic graph model, logistic model, and deep learning model. The exact data requirements differ by model, but what is common is that quiz data are needed by practically all types of KT models.

3. http://pslcdatashop.web.cmu.edu/KDDCup
4. www.assistments.org
5. https://www.khanacademy.org

As shown in Table 3, some models and applications require significantly more none-quiz data. None-quiz data could be classified as student-level and item-level. Student-level data refer to any information that is collects on an individual student, and item-level data refer to any information that is collects on items.

Table 3

Data extracted for each article selected. The numbers in brackets represent the total of studies.

Quiz data	None-quiz data	
Exercise tag, the correctness of responses(48)	Student-level(10)	Students' longitudinal EEG signals, number of skills completed, average response time, and instructional interventions, etc.
	Item-level(8)	Attempt count, first action, and the intrinsic relations of KCs, etc.

4.4 Knowledge Tracing Models

In this section, we review the notable algorithms within the proposed classification frameworks. Instead of an exhaustive list, only the most notable and promising advancements of each category will be reported in this section.

4.4.1 Probabilistic graph model.

Bayesian Knowledge Tracing (BKT), as shown in Fig. 4, uses a 2-node Bayesian network to represent the relations between the observable node Q (the responses of a student) and hidden node K (the knowledge state). The BKT model assumes that, for each skill, the student is either known or unknown. The BKT model typically assumes that forgetting doesn't occur. Additionally, BKT assumes four probability factors for each skill, each of them having a numerical value from 0 to 1:

1. $P(L_0)$, the probability that the KC is already known before the first time to use the skill in problem-solving.
2. P(T), the probability that the KC will be learned at each opportunity to use the skill.
3. P(S), the probability that the student will guess correctly if the KC is not known.
4. P(G), the probability that the student will slip (e.g. make a mistake) if the KC is known.

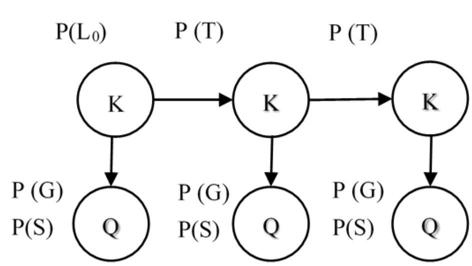


Figure 4

BKT probability graph structure.

Whenever the student has an opportunity to use a KC, the probability that the student knows the KC is updated using formulas derived from Bayes' Theorem (1)-(3). Action_n means the actual correctness of the exercise.

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

$$P(L_{n-1}|Incorrect_n) = \frac{P(L_{n-1})*(P(G))}{P(L_{n-1})*(P(S))+(1-P(L_{n-1}))*(1-P(G))}. \quad (2)$$

$$P(L_n) = P(L_{n-1}|\text{Action}_n) + (1 - P(L_{n-1}|\text{Action}_n)) * P(T) \cdot \quad (3)$$

It has been proved that the BKT parameter space is non-convex(J. Beck & K.-m. Chang, 2007). That is, the parameter estimation task of BKT is subject to local maxima rather than global maxima. This problem of multiple (differing) sets of parameter values that make identical predictions in BKT is known as identifiability (J. Beck & K.-m. Chang, 2007). Besides,if the value of P(G) or P(S) is greater than 0.5, this may cause the model to degenerate (R. S. J. D. Baker et al., 2008). As shown in Table 4, to avoid the above problems, various approaches to parameters fitting for BKT were proposed.

Table 4

Comparisons between different BKT parameter fitting approaches.

Approaches	Fitting Strategies	Pros and Cons
The Baseline Approach (R. S. J. D. Baker et al., 2008)	Each of the four parameters is a value between 0 and 1.	Has the problem of model degeneracy and model identifiability.
The Bounded Guess and Slip Approach (R. S. J. D. Baker et al., 2008; Corbett & Anderson, 1994)	The guessing parameter is bounded to be between 0 and 0.3, and the slip parameter is bounded to be between 0 and 0.1.	Avoids the problem of model degeneracy theoretically but may be inconsistent.
The Dirichlet Priors Approach (R. S. J. D. Baker et al., 2008; Beck, 2007; J. E. Beck & K.-m. Chang, 2007)	A Dirichlet probability distribution is found for how often different values of each parameter are seen based on the student's performance data across multiple skills, and then the parameters of all skills are constrained by these prior probabilities.	Alleviates the problem of model identifiability but have the problem of model degeneracy and the issue of generating the Dirichlet is still a concern.

BKT-Contextual Guess and Slip (CGS) (R. S. J. D. Baker et al., 2008; R. S. J. D. Baker et al., 2010)	Using machine learning to make contextual estimations of P(S) and P(G).	Alleviates the problem of model degeneracy and model identifiability.
BKT-Brute force grid search (BF) (R. S. J. D. Baker et al., 2010; Z. Pardos & N. Heffernan, 2010)	Based on the entire parameter space and by setting boundaries for exhaustive search.	Has the problem of model identifiability and high computational cost.
BKT-Expectation-Maximization (EM) (Beck, 2007; Chang et al., 2006; Z. Pardos & N. Heffernan, 2010; Z. A. Pardos & N. T. Heffernan, 2010; Spaulding & Breazeal, 2015)	Based on the student's performance data to estimate the model parameters by finding parameters that maximize the data likelihood.	Has the problem of model identifiability and local maximum.
BKT-Empirical Probabilities (EP) (Hawkins et al., 2014; Junjie et al., 2014)	Annotating performance data with knowledge.	Non-degenerate but sacrifices the precision.
Clustering parameters across similar skills (Z. A. Pardos et al., 2012; Ritter et al., 2009)	Finding a small number of parameters sets that provide good fits across a wide range of data by clustering, rather than searching a large space of parameters.	Reduces the parameter space.

Probabilistic graph models provide a flexible framework for modeling large collections of variables with complex interactions. Alteration of some basic assumptions and framework of standard BKT yields some models that are more predictive of student performance. Some research tries to add item difficulty to BKT, such as Knowledge Tracing-Item Difficulty Effect Model (KT-IDEM) (Zachary A. Pardos & Heffernan, 2011). KT-IDEM aims to give each question its P(S) and P(G) to effectively extend to capture item difficulty by adding an extra node and arc for each question. Qiu et al. (2011) extended the BKT model to consider forgetting behavior and developed a model called KT-Forget that incorporates the time that has elapsed between opportunities into the BKT model. Yutao and Heffernan (2013) relaxed the assumption of binary correctness by replacing the discrete performance node with a continuous partial credit node and giving a new model named KTPC. Kaser et al. (2014) proposed to use dynamic Bayesian networks (DBNs) to model skill hierarchies within a learning domain, use a log-linear formulation, and apply a constrained optimization to identify the parameters of the DBN.

4.4.2 Logistic regression model.

Performance Factor Analysis (PFA) (Pavlik Jr et al., 2009) uses a logistic regression over aggregated performance to determine students' performance for each skill. PFA defines the probability of success to an item k by a student i as:

$$p(i, j \in KCs, k \in Items, s, f) = \frac{1}{1+e^{-(\beta_k+\sum_{j \in KCs}(\gamma_j s_{i,j}+\rho_j f_{i,j}))}} \quad (4)$$

Where β_k is the easiness of item k, $s_{i,j}$ denotes the number of correctly solved items for student i at KC j, while $f_{i,j}$ denotes the number of prior failures for student i at KC j. The fixed effects γ_j and ρ_j therefore denote the learning rates for successes and failures, respectively. The parameters (β,γ,ρ) could be estimated by maximum likelihood estimation.

Inspired by the logistic regression over the learning and forgetting probabilities, Gonzalez-Brenes et al. (2013) proposed a general method that allows efficient general features (e.g., subskills, problem's difficulty, and student ability etc) into KT model named Feature-Aware Student Knowledge Tracing (FAST). Xu and Mostow (2011) restructured a KT model using logistic regression over each step's subskills to model the learning and forgetting probabilities for overall knowledge required by the step. Zhou et al. (2017) introduced a group of multi-subskill models that integrate all the four parameters (learning rate, forgetting probability, guessing and slipping probability) and item difficulty through logistic regression in the KT models

4.4.3 Deep learning model.

Based on the use of a recurrent neural network, DKT (Piech et al., 2015) is the first model that exhibited promising results using recurrent neural networks and suggested a promising new line of research for KT in deep learning. As shown in Fig.5, the input sequence of the DKT model is described as encoded exercise tags of a student $X_T=(x_1,x_2,\ldots,x_t)$. It undergoes a series of transformations via a hidden layer and forms a sequence of hidden states $(h_1,h_2,\ldots,h_T)$. The output is the probability of getting the exercise corrects:

$$h_t = \tanh(W_{hx}x_t + W_{hh}h_{t+1} + b_h) \quad (5)$$

$$y_t = \sigma(W_{hy}h_t + b_y) \quad (6)$$

In DKT, both tanh and sigmoid functions are applied element-wise and parameterized by an input weight matrix W_{hx}, the recurrent weight matrix W_{hh}, initial state h_0, and readout weight matrix W_{hy}. Biases for latent and readout units are represented by b_h and b_y.

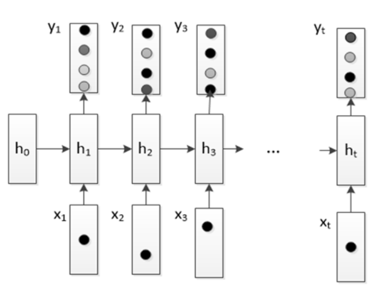


Figure 5
The architecture of deep knowledge tracing.

The objective function of the model is the negative log-likelihood of the observation sequence of student performances under the model and could be minimized using stochastic gradient descent on mini-batches. The objective function is as follows:

$$\mathcal{L} = \frac{1}{\sum_{i=1}^{n}(T_i - 1)} \left(\sum_{i=1}^{n} \sum_{t=1}^{T_i - 1} l\left(y_t^i \cdot \delta\left(q_{t+1}^i\right), a_{t+1}^i\right) \right) \quad (7)$$

where $\ell(\cdot)$ is the binary cross-entropy function, n is the number of students, T_i is the interaction length of student i.

Following the DKT model, there are increasing amounts of researches. L. Zhang et al. (2017) extended the DKT model to incorporated more features at the item-level including first response time, attempt count, and first action. After convert to categorical data, those features were represented as a sparse vector by one-hot encoding as inputs. Then Auto-Encoder was applied to reduce the dimensionality of inputs to DKT. Chen et al. (2018) proposed to incorporate the information of KC structures into the DKT model to solve the problem of model evaluation inaccuracy caused by data sparsity, which specifically refers to considering the pre and post-relationship of KCs. Minn et al. (2018) proposed combines student's learning ability into DKT. K-means was used to clustering the students into a group with similar ability at each time interval first and then combine that information with DKT. Yang and Cheung (2018) designed an automatic system to embed the heterogeneous features implicitly and effectively into the original DKT model.

Besides the recurrent neural network, more and more deep learning algorithms are used for KT modeling. DKVMN (J. Zhang et al., 2017) was proposed to go deeper to trace how specific concepts are mastered by a student based MANN. Casting the knowledge structure as a graph, GKT was proposed by Nakagawa et al. (2019) based on GNN. Lee and Yeung (2019) proposed a new model called Knowledge Query Network (KQN) that uses neural networks to encode student learning activities into knowledge state and skill vectors and model the interactions between the two types of vectors with the dot product. For identifying the relevance between the KCs, Pandey and Karypis (2019) proposed a self-attention based model named Self Attentive Knowledge Tracing (SAKT).

4.5 Performance Metrics of Knowledge Tracing Models.

Towards the two common usages, KT models are usually evaluated by how accurate they predict student's performances, as well as by parameter plausibility (Gong et al., 2011). Parameter plausibility is often tested by comparing them to an external gold standard, and metrics are used to quantifying the quality of predictions. As shown in Figure 6, classification metrics of Area Under the Curve (AUC) and Accuracy are often adopted for the task of evaluating the prediction accuracy of the KT models. Besides, considering the KT task as a regression problem, regression metrics of Root Mean Square Error (RMSE) and Mean Average Error (MAE) are usually used for model performance evaluation, with lower values meaning higher accuracy.

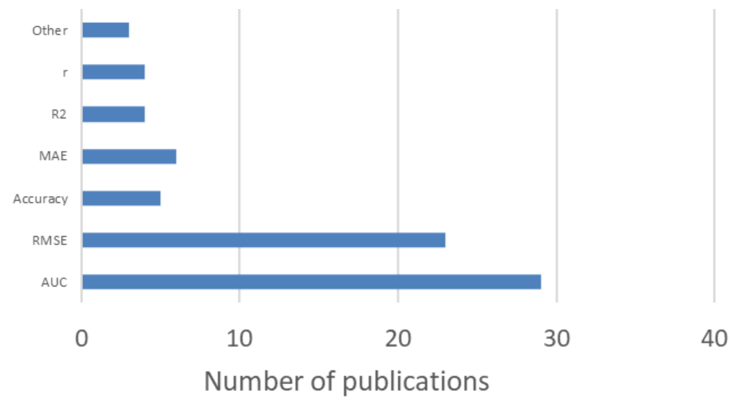


Figure 6
Performance metrics of the selected studies

5. Discussion

5.1 The Influence of Assumptions for Knowledge Tracing Models

Through summarizing the assumptions of existing researches, results in Table 2 indicate that most existing KT models consider only a fragment of the assumptions that relate to KCs, students' personality and their learning process. Educational psychologists have long converged that the knowledge construction procedure of students is not static but constantly evolving (Wang et al., 2013). The problem of optimizing model performance and increasing model reliability in the KT model remains under-explored. Existing work either neglects some fact that affects performance (e.g.forgetting) or assumes it's influence on student knowledge state is constant (e.g. item difficulty), and this is unrealistic in the actual learning process. Future studies should take into account the reasonable and comprehensiveness of the underlying assumptions of what it is that makes the KT model successfully infer students' knowledge state.

5.2 The Results Concluded by Data of Knowledge Tracing Models

Table 3 shows the input data of the KT model found in studies, and the number of related articles. It can be found that quiz data is essential for the KT model to measure students' subject-specific knowledge state. However, the academic performance at a point of time is insufficient to reflect a clear picture of students' learning process (Schrader & Erwin, 1991). Modeling a credible student's profile that reflects the impacts of the learner's characteristics during the learning process can be an interesting research direction for the KT task based on the following evidence: (1) In line with the assumption in cognitive diagnosis (Jiao et al., 2019), process data is worthy of exploration and the integration with quiz data can enhance our evidence base for KT task; (2) 29.2% (14/48) of research on KT modeling analyzed one-quiz data as the supplementary data source and

5.3 Compare Analysis of Knowledge Tracing Models

As for a KT model, estimating accurate and explainable prediction of students' knowledge state requires the combination of well-chosen assumptions, well data collection, and well-implemented algorithms. Unfortunately, there isn't any generalizable model(s) that can be applied in assorted circumstances. The strengths and weaknesses of each kind of KT model are summarized as follows.

According to the prior knowledge of pedagogy and experts, the relations between the observable variable (the responses of a student) and the hidden variable (the knowledge state) could be observed and further hypothesized by various probabilistic graph model (Lafferty et al., 2001; Pirolli & Kairam, 2013). The advantage of the probabilistic graph model is that it can make good use of pedagogical theory and has strong interpretability. In the case of less training data, the model also performs well. However, the performance is highly dependent on the experts' understanding of the scenario, and it cannot explore new KCs that have not been defined by the experts.

The logistic regression model is simple and easy to understand. These models do not describe how a student's knowledge state concerning one skill is affected by another. Instead, it takes a student parameter as the only factor that relates the knowledge state of different skills. Moreover, a skill is explained by the regression coefficient for the skill-specific covariates, from which we cannot tell the structure of the skill domain directly.

The deep learning model mainly uses some existing deep learning algorithms to solve a series of problems in KT. Deep learning has led to important improvements in KT tasks (Ding & Larson, 2019; Long & Pengyu, 2017). While those models need a huge and diverse amount of training data to show good performance. Besides, the quality of the data is also very important. The sparseness of student's exercise data limits the model's performance and application (Vinitra Swamy et al., 2018). Besides, deep learning technology has poor interpretability. In most cases, it can only explain what the model output represents, while the intermediate process is a black box, which cannot explain why such output results are obtained. However, in the KT task, it is hoped that the model can pinpoint intuitively which KCs a student is good at or unfamiliar with. With that in mind, an effective approach to deal with sparse data and well-interpretability is crucial to a deep learning-based KT model.

6. Conclusion and Future Research

To answer the three research questions, we examined the extant literature in an attempt to identify the current state of research in the field of KT modeling. We proposed a coding scheme and summarized KT models from the perspective of assumptions, data, and algorithms. Based on our analysis, some promising future extensions are detailed as follows.

6.1 Perspectives of Assumption and Data

Student learning is influenced by many factors in an authentic learning environment. Heterogeneous data, such as textual and picturial information and student interaction data, needs to be considered in KT modeling. Deep learning has been fruitful in many fields such as natural language processing (Tom et al., 2018), computer vision (Athanasios et al., 2018), and speech recognition (Deng et al., 2013). Such advanced techniques should be introduced to process heterogeneous data and embed heterogeneous data into a high dimensional space to facilitate students' knowledge state. Besides, traditional pedagogical theories, such as memory curve (Gruber, 1992) and forgetting curve (Averell & Heathcote, 2011) should also be considered in the modeling process to strengthen the rationality of model construction and further improve the performance of KT models.

6.2 Perspectives of Algorithms

In general, ensemble modeling is a technique of combining two or more algorithms of similar or dissimilar types called base learners (Tran et al., 2008). This method offers one of the most convincing ways to build highly accurate predictive models that incorporate the predictions from all the base learners. For example, Baker et.al (Ryan S. J. D. Baker et al., 2011) ensembled multiple models for the KT task and found that ensemble models performed comparably to or slightly better than the best individual models in predicting future performance within the tutor software. Said differently, a single KT model based on one data sample can have limits as using specific modeling techniques can present similar drawbacks. By combining different models, ensembled models may help offset those limitations and provide more trustworthy information to instructors and other education stakeholders.

Acoknowledgement

This paper was supported by National Natural Science Foundation of China (NSFC) (No.61807013), Fundamental Research Funds for the Central Universities (No. CCNU20QN028), Special Research Project on Teacher Education (No.CCNUTEIII 2021-07) and Teaching Innovation Research Project (No. ZNXBJY202115).

References

Appleton, K., & Beasley, W. (1994). Students' learning in science lessons: Towards understanding the learning process. *Research in Science Education - RES SCI EDUC*, *24*, 11-20. https://doi.org/10.1007/BF02356324

Athanasios, V., Nikolaos, D., Anastasios, D., & Eftychios, P. (2018). Deep learning for computer vision: A brief review. *Computational Intelligence and Neuroscience*,2018,(2018-2-1), 2018, 1-13.

Averell, L., & Heathcote, A. (2011). The form of the forgetting curve and the fate of memories. *Journal of Mathematical Psychology, 55*, 25-35. https://doi.org/10.1016/j.jmp.2010.08.009

Baker R.S.J.., Corbett A.T., Aleven V. (2008) More accurate student modeling through contextual estimation of slip and guess probabilities in bayesian knowledge tracing. In: Woolf B.P., Aïmeur E., Nkambou R., Lajoie S. (eds) *Intelligent Tutoring Systems. ITS 2008. Lecture Notes in Computer Science,* vol 5091. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-69132-7_44

Baker, R. S. J. D., Corbett, A. T., Gowda, S. M., Wagner, A. Z., MacLaren, B. A., Kauffman, L. R., Mitchell, A. P., & Giguere, S. (2010).Contextual slip and prediction of student performance after use of an intelligent tutor. In: De Bra P., Kobsa A., Chin D. (eds) *User Modeling, Adaptation, and Personalization. UMAP 2010. Lecture Notes in Computer Science,* vol 6075. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-13470-8_7

Baker, R. S. J. D., Pardos, Z. A., Gowda, S. M., Nooraei, B. B., & Heffernan, N. T. (2011). Ensembling predictions of student knowledge within intelligent tutoring systems. *Proceedings of the 19th international conference on User modeling, adaption, and personalization,* Girona, Spain.

Beck, J. (2007). Difficulties in inferring student knowledge from observations (and why you should care) (Eds.). *Educational Data Mining: Supplementary Proceedings of the 13th International Conference of Artificial Intelligence in Education.*

Beck J.E., Chang K. (2007) Identifiability: A Fundamental Problem of Student Modeling. In: Conati C., McCoy K., Paliouras G. (eds) *User Modeling 2007. UM 2007. Lecture Notes in Computer Science,* vol 4511. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-73078-1_17

Chang, K.M., Beck, J., Mostow, J., & Corbett, A. (2006). A Bayes Net Toolkit for student modeling in intelligent tutoring systems. In M. Ikeda, K. D. Ashley, & T. W. Chan (Eds.), *Intelligent Tutoring Systems, Proceedings* (Vol. 4053, pp. 104-113).

Chen, P., Lu, Y., Zheng, V. W., & Pian, Y. (2018). Prerequisite-driven deep knowledge tracing. In *2018 IEEE International Conference on Data Mining (ICDM)* (pp. 39-48). IEEE.

Cheung, L. P., & Yang, H. (2017). Heterogeneous features integration in deep knowledge tracing. In *International Conference on Neural Information Processing* (pp. 653-662). Springer, Cham. https://doi.org/10.1007/978-3-319-70096-0_67

Corbett, A. T., & Anderson, J. R. (1994). Knowledge tracing: Modeling the acquisition of procedural knowledge. *User Modeling and User-Adapted Interaction, 4*(4), 253-278.

Deng, L., Hinton, G., & Kingsbury, B. (2013). New types of deep neural network learning for speech recognition and related applications: an overview (Eds.). *IEEE International Conference on Acoustics*.

Ding, X., & Larson, E. (2019).Why deep knowledge tracing has less depth than anticipated In: *Proceedings of The 12th International Conference on Educational Data Mining (EDM 2019)*, Collin F. Lynch, Agathe Merceron, Michel Desmarais, & Roger Nkambou (eds.) 2019, pp. 282 - 287.

Freedman, R., Ali, S., & McRoy, S. (2000). What is an intelligent tutoring system. *Intelligence, 11*(3), 15-16.

Gong, Y., Beck, J. E., & Heffernan, N. T. (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.

Gonzalez-Brenes, J., Huang, Y., & Brusilovsky, P. (2013). Fast: Feature-aware student knowledge tracing In *Proceedings of NIPS 2013 Workshop on Data Driven Education*.

Gruber, H. (1992). The learning curve in the production of semiconductor memory chips. *Applied Economics, 24*(8), 885-894. https://doi.org/10.1080/00036849200000056

Gu J., Wang Y., & Heffernan N.T. (2014) Personalizing knowledge tracing: should we individualize slip, guess, prior or learn rate?. In: Trausan-Matu S., Boyer K.E., Crosby M., Panourgia K. (eds) *Intelligent Tutoring Systems. ITS 2014. Lecture Notes in Computer Science,* vol 8474. Springer, Cham. https://doi.org/10.1007/978-3-319-07221-0_92

Han, J. W., Jo, J. C., Ji, H. S., & Lim, H. S. (2016). A collaborative recommender system for learning courses considering the relevance of a learner's learning skills. *Cluster Computing, 19*(4), 2273-2284. https://doi.org/10.1007/s10586-016-0670-x

Hawkins W.J., Heffernan N.T., Baker R.S.J.D. (2014) Learning Bayesian knowledge tracing parameters with a knowledge heuristic and empirical probabilities. In: Trausan-Matu S., Boyer K.E., Crosby M., Panourgia K. (eds) *Intelligent Tutoring Systems. ITS 2014. Lecture Notes in Computer Science,* vol 8474. Springer, Cham. https://doi.org/10.1007/978-3-319-07221-0_18

Hilles, M. M., & Naser, S. S. A. (2017). Knowledge-based Intelligent Tutoring System for Teaching Mongo Database. *European Academiic Research,* Vol IV, Issue 10

Jiao, H., Liao, D., & Zhan, P. (2019). Utilizing process data for cognitive diagnosis. *Handbook of Diagnostic Classification Models* (pp. 421-436). Springer.

Jin, Z., Ma, K., Liu, K., & Ji, K. (2019). Exercises recommendation in adaptive learning system. In *Proceedings of the 2nd International Conference on Big Data Technologies* (pp. 97-100).

Käser T., Klingler S., Schwing A.G., Gross M. (2014) Beyond knowledge tracing: Modeling skill topologies with Bayesian networks. In: Trausan-Matu S., Boyer K.E., Crosby M., Panourgia K. (eds) *Intelligent Tutoring Systems. ITS 2014. Lecture Notes in Computer Science,* vol 8474. Springer, Cham. https://doi.org/10.1007/978-3-319-07221-0_23

Kasurinen, J., & Nikula, U. (2009). Estimating programming knowledge with Bayesian knowledge tracing. *Proceedings of the 14th annual ACM SIGCSE conference on Innovation and technology in computer science education*, Paris, France.

Lafferty, J., McCallum, A., & Pereira, F. C. (2001). Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In *Proceedings of the Eighteenth International Conference on Machine LearningJune* 2001 (pp 282–289)

Lee, J., & Yeung, D.-Y. (2019). Knowledge query network for knowledge tracing: How knowledge interacts with skills. *Proceedings of the 9th International Conference on Learning Analytics & Knowledge*, Tempe, AZ, USA.

Minn, S., Yu, Y., Desmarais, M.C., Zhu, F., & Vie, J. (2018). Deep knowledge tracing and dynamic student classification for knowledge tracing. *2018 IEEE International Conference on Data Mining (ICDM),* 1182-1187.

Mongkhonvanit, K., Kanopka, K., & Lang, D. (2019). Deep knowledge tracing and engagement with MOOCs. *Proceedings of the 9th International Conference on Learning Analytics & Knowledge,* Tempe, AZ, USA.

Nakagawa, H., Iwasawa, Y., & Matsuo, Y. (2019). Graph-based knowledge tracing: Modeling student proficiency using graph neural network. *IEEE/WIC/ACM International Conference on Web Intelligence,* Thessaloniki, Greece.

Pandey, S. & Karypis, G. (2019). A self attentive model for knowledge tracing. In *Proceedings of the 12th International Conference on Educational Data Mining*.

Pardos, Z. A., & Heffernan, N. T. (2010). Navigating the parameter space of Bayesian Knowledge Tracing models: Visualizations of the convergence of the Expectation Maximization algorithm. *Educational Data Mining* 2010.

Pardos Z.A., Heffernan N.T. (2010) Modeling individualization in a Bayesian networks implementation of knowledge tracing. In: De Bra P., Kobsa A., Chin D. (eds) *User Modeling, Adaptation, and Personalization. UMAP 2010. Lecture Notes in Computer Science*, vol 6075. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-13470-8_24

Pardos, Z. A., & Heffernan, N. T. (2011). KT-

IDEM: Introducing item difficulty to the knowledge tracing model. *Proceedings of the 19th international conference on User modeling, adaption, and personalization.*

Pardos Z.A., Trivedi S., Heffernan N.T.& Sárközy G.N. (2012) Clustered knowledge tracing. In: Cerri S.A., Clancey W.J., Papadourakis G., Panourgia K. (eds) *Intelligent Tutoring Systems. ITS 2012. Lecture Notes in Computer Science,* vol 7315. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-30950-2_52

Pavlik Jr, P., Cen, H., & Koedinger, K. (2009). Performance factors analysis - A new alternative to knowledge tracing. In *Proceedings of the 2009 conference on Artificial Intelligence in Education: Building Learning Systems that Care: From Knowledge Representation to Affective Modelling*(pp. 531-538).

Piech, C., Bassen, J., Huang, J., Ganguli, S., Sahami, M., Guibas, L., & Sohl-Dickstein, J. (2015). Deep knowledge tracing. (Eds.). *Proceedings of the 28th International Conference on Neural Information Processing Systems* - Volume 1.

Pirolli, P., & Kairam, S. (2013). A knowledge-tracing model of learning from a social tagging system. *User Modeling and User-Adapted Interaction, 23*(2-3), 139-168. https://doi.org/10.1007/s11257-012-9132-1

Qiu, Y., Qi, Y., Lu, H., Pardos, Z. A., & Heffernan, N. T. (2011).Does time matter? Modeling the effect of time with Bayesian knowledge tracing. *EDM.*

Ritter, S., Harris, T. K., Nixon, T., Dickison, D., Murray, R. C., & Towle, B. (2009). Reducing the knowledge tracing space. *International Working Group on Educational Data Mining.*

Schodde, T., Bergmann, K., & Kopp, S. (2017). Adaptive robot language tutoring based on bayesian knowledge tracing and predictive decision-making. *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction,* Vienna, Austria.

Schrader, D. E., & Erwin, T. D. (1991). Assessing student learning and development: A guide to the principles, goals, and methods of determining college outcomes. *Journal of Higher Education,* 63(4), 463.

Self, J. A. (1990). Bypassing the intractable problem of student modelling. *Intelligent tutoring systems: At the crossroads of artificial intelligence and education, 41*, 1-26.

Sha L., & Hong P. (2017) Neural knowledge tracing. In: Frasson C., Kostopoulos G. (eds) *Brain Function Assessment in Learning. BFAL 2017. Lecture Notes in Computer Science*, vol 10512. Springer, Cham. https://doi.org/10.1007/978-3-319-67615-9_10

Spaulding, S., & Breazeal, C. (2015, March). Affect and inference in Bayesian knowledge tracing with a robot tutor. In *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction Extended Abstracts,*Portland, Oregon, USA.

Sun, X., Zhao, X., Ma, Y., Yuan, X., He, F., & Feng, J. (2019, May). Muti-behavior features based knowledge tracking using decision tree improved DKVMN. In *Proceedings of the ACM Turing Celebration Conference-China* (pp. 1-6).

Swamy, V., Guo, A., Lau, S., Wu, W., Wu, M., Pardos, Z., & Culler, D. (2018) Deep knowledge tracing for free-form student code progression. In: Penstein Rosé C. et al. (eds) *Artificial Intelligence in Education. AIED 2018. Lecture Notes in Computer Science,* vol 10948. Springer, Cham. https://doi.org/10.1007/978-3-319-93846-2_65

Tran, L. M., Rizk, M. L., & Liao, J. C. (2008). Ensemble modeling of metabolic networks. *Biophysical journal*, *95*(12), 5606-5617.

Vlahavas, I. P., & Spyropoulos, C. D. (2002). Methods and applications of artificial intelligence (Eds.). *Second Hellenic Conference on AI*, SETN.

Wang, X., Berger, J., & Burdick, D. (2013). Bayesian analysis of dynamic item response models in educational testing. *The Annals of Applied Statistics*, 7. https://doi.org/10.1214/12-AOAS608

Wang Y., & Heffernan N. (2013) Extending knowledge tracing to allow partial credit: Using continuous versus binary nodes. In: Lane H.C., Yacef K., Mostow J., Pavlik P. (eds) *Artificial Intelligence in Education. AIED 2013. Lecture Notes in Computer Science,* vol 7926. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-39112-5_19

Xu, Y., Chang, K.-M., Yuan, Y., & Mostow, J. (2014). Using EEG in knowledge tracing. (Eds.). *Educational Data Mining* 2014.

Xu, Y., & Mostow, J. (2011).Using logistic regression to trace multiple sub-skills in a dynamic Bayes net. (Eds.). *EDM*.

Yang, H., & Cheung, L. P. (2018). Implicit heterogeneous features embedding in deep knowledge tracing. *Cognitive Computation, 10*(1), 3-14. https://doi.org/10.1007/s12559-017-9522-0

Young, T., Hazarika, D., Poria, S., & Cambria, E. (2018). Recent trends in deep learning based natural language processing. i*eee Computational intelligence magazine, 13*(3), 55-75.

Yudelson, M. V., Medvedeva, O. P., & Crowley, R. S. (2008). A multifactor approach to student model evaluation. *User Modeling and User-Adapted Interaction, 18*(4), 349-382. https://doi.org/10.1007/s11257-007-9046-5

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

Zhang, L., Xiong, X., Zhao, S., Botelho, A., & Heffernan, N. T. (2017). Incorporating rich features into deep knowledge tracing. *Proceedings of the Fourth (2017) ACM Conference on Learning at Scale,* Cambridge, Massachusetts, USA.

Zhou, X., Wu, W., & Han, Y. (2017). Modeling multiple subskills by extending knowledge tracing model using logistic regression. In *2017 IEEE International Conference on Big Data (Big Data)* (pp. 3994-4003). IEEE.

Zhu, J., Zang, Y., Qiu, H., & Zhou, T. (2018). Integrating temporal information into knowledge tracing: A temporal difference approach. *IEEE Access,* 6, 27302-27312. https://doi.org/10.1109/ACCESS.2018.2833874