Proceedings of the 5th International Conference on Computing and Data Science
DOI: 10.54254/2755-2721/16/20230885

The study and analysis related to the evolution of the knowledge tracing research

Yu Bai

The department of computer science, Hefei University of Technology, Xuan Cheng, 242000, China

yuyy@xy.hfut.edu.cn

Abstract. In the field of education, the level of awareness of students' knowledge status largely affects students' learning efficiency. Knowledge Tracing, as a method of modeling students' learning status, obtaining their knowledge status to predict their future learning can provide more targeted learning solutions, which is extremely useful for improving students' learning efficiency and implementing tailor-made education. This paper introduces the related research methods and their core ideas in the order of probability-based models, logistic function-based models, and deep learning-based models by investigating the development of this field since its introduction. Among the probability-based models, the mainstream methods include BKT and DBKT, which are based on Hidden Markov Models with certain assumptions and are not ideal in practical application scenarios. In contrast, logistic-based models consider a variety of factors in the student learning process and have achieved good results. After that, deep learning-based model has achieved excellent assessment results through its powerful feature extraction ability, and many variants have been derived. This paper introduces DKT in detail and briefly introduces its variant models. At last, this paper summarizes some of the mainstream evaluation criteria in this field and gives the widely used datasets and their categories. Finally, this paper proposes some suggestions for the future development of the field in view of the shortcomings of the current research status for reference.

Keywords: knowledge tracing, BKT, LFA, DKT.

1. Introduction

Presently, although the impact of the COVID-19 epidemic on human beings is dissipating, it is indisputable that the epidemic has instigated numerous alterations in recent years. The sphere of education, being one of the most profoundly affected domains, has led to the expeditious advancement of online learning and open learning. It is precisely the massive data generated by this opportunity that Knowledge Tracing (KT) is once again known as a hot research field. As a widely used technology in educational data mining, KT aims to build a model to quantify students’ level of mastery of knowledge over time. It is widely used to predict their future performance based on previous performance, thereby facilitating personalized learning experiences tailored to individual students' specific needs and aptitudes.

Research on Knowledge Tracing can date back to the 1990s, when Corbett and Anderson proposed the first Knowledge Tracing model [1], which used Bayesian networks to analyse students' performance in answering questions to predict their mastery of specific knowledge concepts. After that, the

DOI: 10.54254/2755-2721/16/20230885

introduction of Item Response Theory (IRT) subsequently had a major impact on this area of research, taking more factors into account when modeling students' learning states, and more models based on this theory were subsequently generated with good results. However, with the maturity of computer technology and the emergence of deep learning techniques, Deep Knowledge Tracing (DKT) models and their variants have breathed new life into the field.

Knowledge Tracing holds great research value and significance due to its potential to boost the efficiency of the process of human learning. Thus, it is imperative to conduct comprehensive reviews and summaries of the research findings in this field. However, most of the review articles on this field analyse and compare the classical models in Knowledge Tracing from one aspect, and there is a lack of review studies that cover a wide range of topics and are relatively new in technical aspects. Therefore, this article is divided into several parts to provide a detailed introduction to the development and recent research status of this field. In the section 2, this article firstly introduces what KT is. After that, this paper introduced the current research status of Knowledge Tracing and the subdivision of this research field. Subsequently, a brief introduction to the core ideas of these methods is provided. In the third section, the relevant datasets and some evaluation criteria of this field are presented. This paper also lists some developments and challenges about this field in the future and made a conclusion at last.

2. Basic knowledge tracing models

2.1. Overview of the knowledge tracing

As the flowchart shown in Figure 1, in order to obtain students' learning status, KT assume that there is a student S and a group of exercises E in a learning system. The learning sequence of the student can be described as follows.

$$\mathrm{X} = \{(\mathrm{e}_1, \mathrm{k}_1, \mathrm{r}_1), (\mathrm{e}_2, \mathrm{k}_2, \mathrm{r}_2), (\mathrm{e}_3, \mathrm{k}_3, \mathrm{r}_3), \ldots \ldots, (\mathrm{e}_\mathrm{n}, \mathrm{k}_\mathrm{n}, \mathrm{r}_\mathrm{n})\} \tag{1}$$

where (e_1, k_1, r_1) represents the student's learning interaction record in a certain process, and e_i is a symbol of the exercise completed by them, k_i means the knowledge concepts involved in completing this exercise, r_i represents the score for the exercise. Through the constructed model, it is possible to quantify students' learning data, predict their mastery of internal and implicit knowledge concepts, and predict their performance in unknown exercises. Based on these predictive outcomes, it is feasible to develop personalized learning plans for students, with the objective of enhancing their learning efficiency.

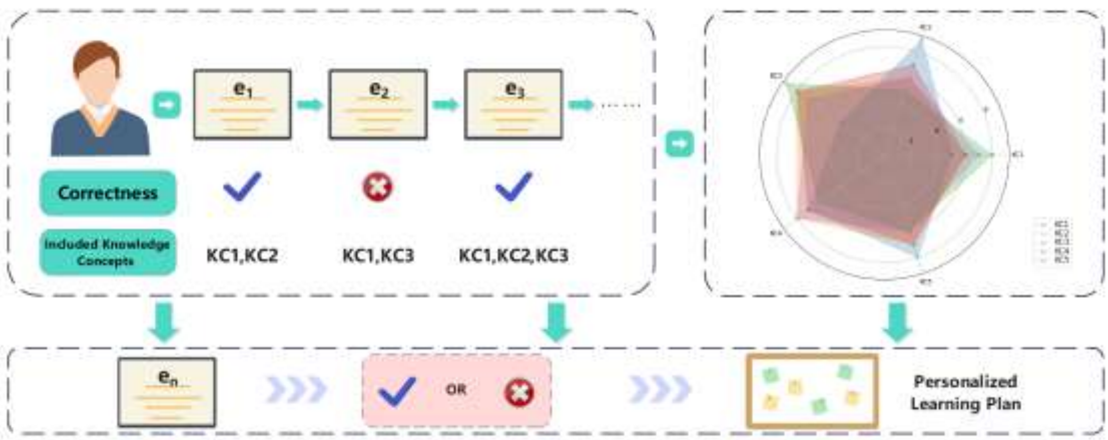

Figure 1. Knowledge Tracing Flowchart.

Since the proposal of the field of knowledge tracing, significant progress has been made in the development of knowledge tracing models. Currently, lots of researchers thinks this field should be divided into three types. They are Bayesian Knowledge Tracing (BKT) based on Hidden Markov

DOI: 10.54254/2755-2721/16/20230885

Models, Logistic Models based on logistic function, and Deep Knowledge Tracing (DKT) which is based on Recurrent Neural Network (RNN), as well as several variations of these models based on these theories. The taxonomy of these three types of models and their subcategories can be summarized as Figure 2 [2].

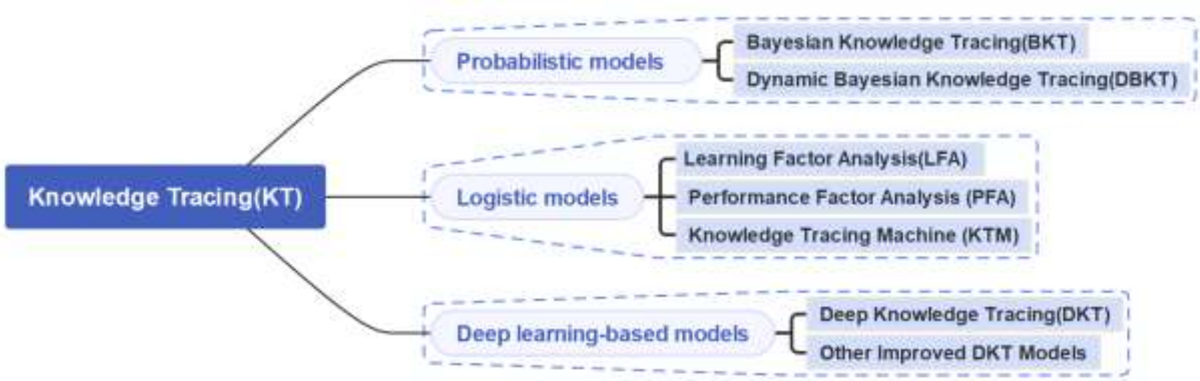

Figure 2. The taxonomy of three types of KT models.

2.2. Probabilistic models

Probabilistic based knowledge tracing model is a common method for assessing students' knowledge levels. It uses probability models such as Bayesian Networks to predict students' mastery of specific knowledge concepts. In probability-based KT, there are two main categories: Bayesian Knowledge Tracing (BKT) and another method based on BKT is Dynamic Bayesian Knowledge Tracing (DBKT). These two methods use Markov model to express a student's learning process. The fundamental notion underlying of them is to use a student's historical answer data to build a Bayesian network model and predict their mastery of future knowledge concepts by continuously updating the model parameters [3].

2.2.1. Bayesian knowledge tracing (BKT)

BKT, as the first model, was proposed by Corbett et al. [1]. In this model, a hidden variable is proposed which is about the student's knowledge state, and it is represented by a binary group (i.e. mastered the knowledge concept, did not master the knowledge concept). The whole model structure is actually an HMM model, and the student's learning process, which can be regarded as a sequential process of HMM, predicts the next state according to the state transfer matrix, and the student's answer result according to the current state.

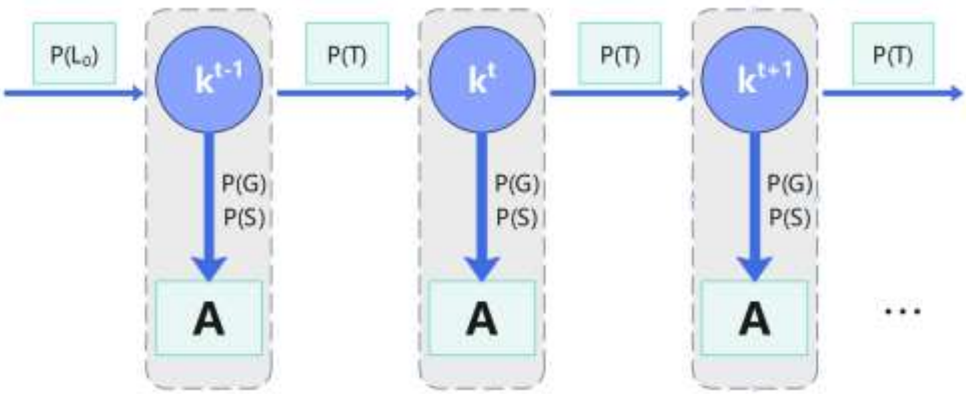

Figure 3. Schematic diagram of BKT model.

DOI: 10.54254/2755-2721/16/20230885

As shown in Figure 3, the modeling using this method involves four parameters, which are the probability of user's initial mastery of knowledge concept k $p(L_0)^k$, the probability of user's mastery through learning $p(T)^k$. The others are the probability of user's mastery of knowledge concept k but wrong answer $p(S)^k$, and the probability that the user does not master the knowledge concept but answers the question correctly $p(G)^k$. With these four parameters, a model of HMM can be constructed. After solving for the corresponding parameters of the model, the mastery of student u at moment $t+1$ for knowledge concept k can be expressed by the following equation:

$$\mathrm{p}(\mathrm{L}_{\mathrm{t+1}})_{\mathrm{u}}^{\mathrm{k}} = \mathrm{p}(\mathrm{L}_{\mathrm{t+1}} \mid \mathrm{Q})_{\mathrm{u}}^{\mathrm{k}} + \left(1 - \mathrm{p}(\mathrm{L}_{\mathrm{t+1}} \mid \mathrm{Q})_{\mathrm{u}}^{\mathrm{k}}\right) \cdot \mathrm{p}(\mathrm{T})^{\mathrm{k}} \tag{2}$$

The probability of student u answering question Q correctly at moment t+1 can then be calculated like this.

$$\mathrm{p}(\mathrm{Q}_{\mathrm{t+1}})_{\mathrm{u}}^{\mathrm{k}} = \mathrm{p}(\mathrm{L}_{\mathrm{t}})_{\mathrm{u}}^{\mathrm{k}} \cdot \left(1 - \mathrm{p}(\mathrm{S})^{\mathrm{k}}\right) + \left(1 - \mathrm{p}(\mathrm{L}_{\mathrm{t}})_{\mathrm{u}}^{\mathrm{k}}\right) \cdot \mathrm{p}(\mathrm{G})^{\mathrm{k}} \tag{3}$$

Although the method is representative in the KT domain and provides real-time feedback on the user's probability of mastery of knowledge concepts, while taking the user's a prior mastery of knowledge concepts into account. However, it is very noteworthy that the model is based on the three premises that the transfer probability $P(T)$ remains constant at different moments, each exercise is independent of each other, and the learning state of students cannot be transferred in the reverse direction (students will not forget the learned knowledge concepts). These assumptions lead to the practical application of the model not being very satisfactory.

2.2.2. Dynamic bayesian knowledge tracing (DBKT)

Since the BKT model cannot realistically reflect a student's learning process, Käser et al. proposed a model in 2017 that can consider the relationships between the required knowledge concepts for each question, called Dynamic Bayesian Knowledge Tracing (DBKT) [4]. Which is based on a Dynamic Bayesian Network and can model the prerequisites, hierarchy, and relationships between knowledge concepts.

Although the mastery of knowledge by students in DBKT is also represented by binary latent variables compared to BKT. However, DBKT introduces $p(F)$ as the forgetting probability of students for knowledge concepts based on the four model parameters of BKT, and also considers the correlations among these KCs. Assuming that X is used to denote the latent space and O to denote the observed space. So $X \times O \to R^d$ can be considered as a process from two spaces: latent space and observed space, to some d-dimensional feature vectors. Then if use c as a normalization constant and use ω as a log-linear model to denote the dependency weights between different KCs, the goal of this model is to find a model parameter $\{p(L_0), p(T), p(S), p(G), p(F), \mathbf{w}\}$ which can maximize the following equation.

$$L(\mathbf{w} = \sum_i \ln\left(\sum_{x_i} \exp(\mathbf{w}^\top f(x_i, y_i) - c)\right) \tag{4}$$

The equation (4) means the likelihood of the joint probability. Where $x_i \in X, y_i \in O$, in this way it is able to represent the association between different KCs, which solves the shortcoming of BKT just modeling for a single knowledge concept and can obtain a better result.

2.3. Logistic models

Logistic models, as a new class, lots of them are based on logistic functions compared to probabilistic models. The basic principle of this approach involves taking into account a wider range of factors that influence students' learning processes. By utilizing logistic function, it becomes possible to predict the probability of a student's mastery. This section will introduce some representative methods in this field.

The first one is Learning Factor Analysis (LFA), the second one is Performance Factor Analysis (PFA) and Knowledge Tracing Machine (KTM) as last.

2.3.1. Learning factor analysis (LFA)

Considering that students' ability keeps changing with the time of learning, LFA has taken into account the initial learning status [5], the difficulty of KCs, and the learning rate of students and expressed them as $\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}$ respectively, so that LFA can be extracted by the formula shown below.

$$p(\theta) = \sigma\left(\sum_{i \in N} \alpha_i S_i + \sum_{j \in KCs} \left(\beta_j + \gamma_j T_j\right) K_j\right) \tag{5}$$

In the above equation,S_i means the covariate for a student i, T_j is used to indicate the covariate of those number of interactions of a student for KC_j. Besides, K_j is used as the covariate of KC_j, and the result of the calculation in parentheses is passed through the sigmoid function to obtain the probability that a student will answer correctly when comes a question. After that, the Additive Factor Model (AFM) was proposed based on LFA, which considered the student's ability, the difficulty of the knowledge in each test and the amount of knowledge obtained by the student in each attempt and had a better result.

2.3.2. Performance factor analysis (PFA)

Performance Factor Analysis [6], as one of the Logistic models, is actually an extension of the LFA, and its core content can be expressed by the following equation:

$$p(\theta) = \sigma\left(\sum_{j \in KCs} \left(\beta_j + \mu_j s_{ij} + \nu_j f_{ij}\right)\right) \tag{6}$$

The meaning of each parameter in this equation is shown in Table 1.

Table 1. Parameters and Meanings of PFA.

Parameter Name	Meaning
f	Prior failures for the KC of the student
s	Prior successes for the KC of the student
β	Easiness of different KCs
μ, ν	Coefficients for s and f denote the learning rates for successes and failures

The parameters and their meanings shown in the above table show that compared to LFA, PFA adds the influence on the success and failure of students in the past attempts with different KCs.

2.3.3. Knowledge tracing machines (KTM)

In 2019, a factorization machine (FM)-based approach was proposed by Vie and named Knowledge Tracing Machines (KTM) [7]. The main idea of this model is to do feature interaction through FM, denoting the total number of features by N. The features contain exercises, knowledge concepts, some factors of learning process (such as previous success or failure information), and some other factors. KTM models the probability of the binary outcome of a student doing a question (correct or incorrect) which is based on a sparse set. The set contains all the features involved in the event, and its prediction is based on the following equation:

$$\psi\left(p(\boldsymbol{x})\right) = \mu + \sum_{k=1}^{N} w_k x_k + \sum_{1 \le k < l \le N} x_k x_l \langle \boldsymbol{v_k}, \boldsymbol{v_l} \rangle \tag{7}$$

The FM method assigns a latent weight vector to each feature and generates the weight vector for feature interactions by taking the dot product of the latent weight vectors. This approach effectively addresses the problem of sparsity in feature interactions which makes it difficult to train models. By applying the widely used FM method in recommendation systems to knowledge tracing, the proposed model achieves a promising result.

2.4. Deep learning-based models

The accuracy of research in KT field is hindered by the fact that probabilistic and logic-based modeling often requires more factors to be considered. With the emergence of deep learning, although its interpretability is controversial, it is widely used in the field of knowledge tracing due to its powerful feature extraction ability, which has injected new vitality into the field. This section will focus on Deep Knowledge Tracing (DKT) and briefly introduce some other Deep learning-based models.

In 2015, Piech et al. first proposed to apply deep learning to knowledge tracing and model the learning process of students using RNN [8]. RNN, as a time-series model, is able to use information from history to make predictions about the future and has good performance on sequential prediction problems.

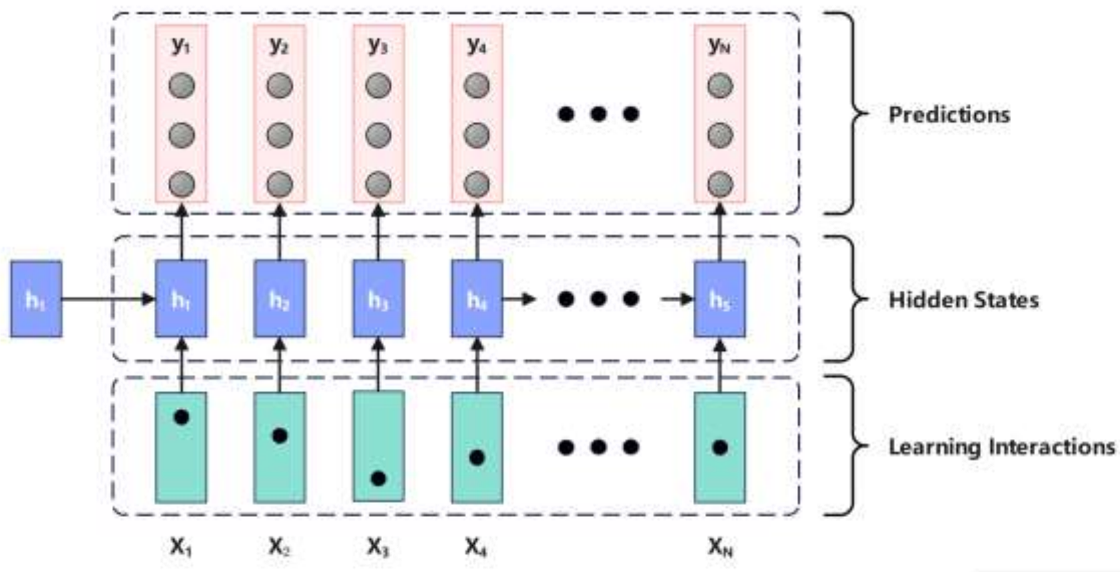

Figure 4. DKT schematic diagram.

DKT is based on RNN, as a seq2seq recurrent neural network model, where the output at each moment is a prediction of the learner's performance at the next moment. The model is divided into three parts, first one is input layer, after that a hidden layer is used, and an output layer to give an output. As shown in the Figure 4, where the input layer is the student's past learning records, the records $\{x_1, x_2, \ldots, x_t\}$ are converted into vector form and as input to the model, through the hidden layer as the memory unit of the model (where h_i represents the student's knowledge state), a sigmoid linear activation layer is used to output the future student performance $\{y_1, y_2, \ldots, y_t\}$, which represents the predicted probability of correctly answering the question.

Through several tests on several benchmark datasets, the model achieves SOTA results, which is largely due to the model's time-series nature, which can record students' knowledge over a longer period of time based on their recent answers and can automatically capture the association between similar topics. However, the model is not interpretable and suffers from both long-term dependency problems and lack of learning features. Therefore, after the DKT model was proposed, many other deep learning-based models in this field were proposed one after another to make up for these points.

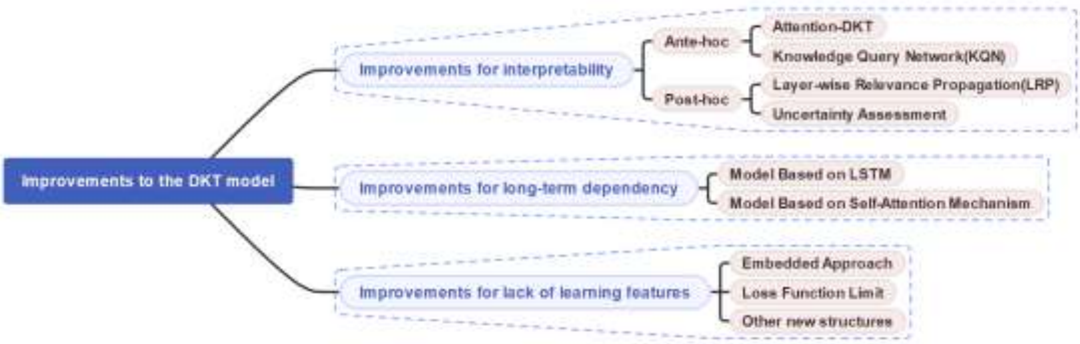

Figure 5. Improvements to the DKT model.

As shown in Figure 5, different types of corresponding solutions are proposed in order to compensate the shortcomings of the DKT model. For interpretability, there are two main categories, Ante-hoc and Post-hoc. Representative models include Attention-DKT, Knowledge Query Network (KQN), Layer-wise Relevance Propagation (LRP) and Uncertainty Assessment. In contrast, Model Based on LSTM or Self-Attention Mechanism solved the long-term dependency to some extent [9]. For the purpose of solve the problem of lack of learning features, Embedded Approach, Loss Function Limit and some other methods are proposed.

3. Evaluation and discussion

In Section 2, several methods are introduced, including probability-based, logistic model-based, and deep learning-based methods. This section will introduce the more general evaluation criteria and related datasets in this field, and the future prospects of the KT field.

3.1. Evaluation criteria and datasets

In knowledge tracking, the commonly used metrics for evaluating models are mainly root mean square error (RMSE) and mean absolute error (MAE) as well as accuracy (ACC) [10]. Sometimes, the area under curve (AUC) of receiver operating characteristic (ROC) can also be used. In these four methods, the first two are based on the regression perspective and the latter two on the classification perspective, and the formulae for calculating these four are shown below:

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{N}(h(x^{(i)}) - y^{(i)})^2} \tag{8}$$

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{N}\left|h\left(x^{(i)}\right) - y^{(i)}\right| \tag{9}$$

$$\mathrm{ACC} = \frac{r}{N} \tag{10}$$

$$\mathrm{AUC} = \frac{\sum_{i\in positive}\mathrm{Enk}_i - \frac{m(1+m)}{2}}{m\times n} \tag{11}$$

In the above equations, N denotes the number of practice interaction records of students, $h\left(x^{(i)}\right)$ is the predicted value of the score of the ith interaction record, $y^{(i)}$ is the true score value, r is the number of correct predictions, m is the number of positive examples, and n is the number of negative examples.

With the rise of educational software and the increasing amount of related data, the dataset on KT has been developed in a more mature way until now, and the following table lists the main datasets currently used in this field.

Table 2. Related Datasets.

Datasets	Subjects and Stages	Datasets	Subjects and Stages	Datasets	Subjects and Stages
ASSIST2009	mathematics (middle school)	ASSISTchall	/	KDDCup2010	/
ASSIST2012	mathematics (middle school)	Statics2011	engineering (university)	Synthetic	virtual students
ASSIST2015	mathematics (middle school)	Junyi	mathematics (primary to high school)	EdNet	English

3.2. *Future prospects*

By modeling students' learning process, knowledge tracking can help students and teachers better understand the learning status and provide more targeted teaching. Today, after decades of development, the field of knowledge tracking has brought certain results and has been widely used in various online learning platforms. However, it is undeniable that the field still has a lot of areas worth exploring, and for the future development of the field, this paper considers the following aspects:

- Propose models with interpretability. The existing models have a good measurement effect, but the interpretability is not considered as an indicator in the measurement effect, which leads to a large number of models based on deep learning technology have achieved good results, but they do not have good interpretability, so this paper believes that this direction is the focus of future research.
- Cold start problem. In the modeling of student learning process, a large amount of data is needed as support, but for the initial state, there is not enough data as support will lead to the system has a cold start problem, which needs to be solved.
- The level of student interaction. In the student learning process, although knowledge tracking can help students achieve personalized learning to a certain extent, it still faces problems such as too little interaction between students and the system and low participation rate. To address this issue, more incentives need to be provided to increase student interaction and participation rates.

4. Conclusion

Knowledge Tracing, as a method to model learners' learning status, can assist students in understanding of their mastery of knowledge and enable instructors to provide more personalized teaching solutions to help them have a better understanding of the knowledge. This paper takes a macroscopic view of the current mainstream methods in this field, which is divided into three classes, probability-based, logistic-based and deep learning-based models. Among the probability-based models, the BKT model and the DBKT based on it are introduced. These models are based on Hidden Markov Models, but these methods are not in line with the actual situation because of lots of prerequisites. In contrast, logistic-based models take into account various characteristics of students' learning process and use logistic function to model, which has achieved good results. With the development of deep learning in these years, through its powerful feature extraction ability and massive of data was recorded, deep learning-based model has been proposed and achieved a SOTA result. However, it is still necessary to face the fact that the use of this method does not good at interpretability and there are some other shortcomings that still need to be solved.

References

[1] Corbett A T Anderson J R 1994 User modeling and user-adapted interaction (Modeling the acquisition of procedural knowledge) p 253-278

DOI: 10.54254/2755-2721/16/20230885

[2] Abdelrahman G Wang Q Nunes B 2023 Knowledge Tracing: A Survey (ACM Computing Surveys) p 1-37
[3] Liu Q Shen S Huang Z et al 2021 A survey of knowledge tracing (arXiv preprint arXiv:210515106)
[4] Käser T Klingler S Schwing a G Gross M 2017 Dynamic Bayesian networks for student modelling (IEEE Transactions on Learning Technologies) p 450-462
[5] Cen H Koedinger K & Junker B 2006 Learning factors analysis–a general method for cognitive model evaluation and improvement (Intelligent Tutoring Systems: 8th International Conference) p 164-175
[6] Pavlik Jr P I Cen H & Koedinger K R 2009 Performance Factors Analysis--A New Alternative to Knowledge Tracing (Online Submission)
[7] Vie J J & Kashima H 2019 Knowledge tracing machines: Factorization machines for knowledge tracing (Proceedings of the AAAI Conference on Artificial Intelligence) p 750-757
[8] Piech C Bassen J Huang J Ganguli S Sahami M Guibas L J & Sohl-Dickstein J 2015 Deep knowledge tracing (Advances in neural information processing systems)
[9] [Pandey S Karypis G 2019 A self-attentive model for knowledge tracing (arXiv preprint arXiv:190706837)
[10] Dalianis H 2018 Evaluation metrics and evaluation Clinical text mining: secondary use of electronic patient records 45-53