Intervention-BKT: Incorporating Instructional Interventions into Bayesian Knowledge Tracing

Chen Lin and Min Chi

The Department of Computer Science,
North Carolina State University, USA
{clin12,mchi}@ncsu.edu
http://www.csc.ncsu.edu

Abstract. Bayesian Knowledge Tracing (BKT) is one of the most widely adopted student modeling methods in Intelligent Tutoring Systems (ITSs). Conventional BKT mainly leverages sequences of observations (e.g. *correct*, *incorrect*) from student-system interaction log files to infer student latent knowledge states (e.g. *unlearned*, *learned*). However, the model does not take into account the instructional interventions that generate those observations. On the other hand, we hypothesized that various types of instructional interventions can impact student's latent states differently. Therefore, we proposed a new student model called Intervention-Bayesian Knowledge Tracing (Intervention-BKT). Our results showed the new model outperforms conventional BKT and two factor analysis based alternatives: Additive Factor Model (AFM) and Instructional Factor Model (IFM); moreover, the learned parameters of Intervention-BKT can recommend adaptive pedagogical policies.

Keywords: Knowledge Tracing, Hidden Markov Model, Input Output Hidden Markov Model, Student Modeling, Instructional Intervention

1 Introduction

Bayesian Knowledge Tracing (BKT) is one of the most widely adopted student modeling methods. BKT leverages sequences of observations (e.g. *correct*, *incorrect*) from student-system interaction log files to continually update the estimate of student latent knowledge (e.g. *unlearned*, *learned*), regardless of the instructional interventions generate the corresponding observations. Instructional interventions are actions initiated by the system guiding student learning activity. For example, two common instructional interventions are *elicit* and *tell*: *Elicit* represents asking a student what is the next step, while *tell* means delivering educational content via a written statement that reveals the next step.

While conventional BKT does not take into account various types of instructional interventions that generate student observations, they can directly impact student's latent states differently. For example, a correct observation should be treated differently depending on whether it is drawn from an *open-ended question* or a *multiple choice* one. Similarly, a correct observation should be treated differently depending on whether it is generated by the student (e.g., the tutor *elicits*) or the tutor (e.g., the tutor *tells*). We proposed a new approach Intervention-Bayesian Knowledge Tracing (Intervention-BKT), which can: 1) incorporate different types of instructional interventions into student model, and 2) can tease apart their effects on student's performance by training a separate set of parameters for each intervention type.

Much of the prior research on evaluating student models did not take different pedagogical strategies that an ITS can employ into account. In our experiment, we trained our models on datasets generated from four training corpus following different pedagogical strategies that vary from ineffective policies to effective ones. We investigated whether Intervention-BKT would outperform conventional BKT regardless of the pedagogical strategies employed. Additionally, we extended our comparison to two other factor analysis based approaches: the widely applied Additive Factor Model (AFM) [4] and the Instructional Factor Model (IFM) [5]. The latter can be seen as an extension of AFM to incorporate different instructional interventions. Finally, we use the parameters learned from our Intervention-BKT models to design adaptive and personalized pedagogical policies.

2 Related Work

In recent years a variety of extensions of BKT have been investigated. Pardos and Heffernan[10] proposed KT-IDEM model by adding a problem difficulty node to the conventional BKT model. Their results showed that KT-IDEM model significantly outperformed BKT on ASSISTments dataset but not on Cognitive Tutor dataset[10]. While KT-IDEM assumes that S_t only depends on S_{t-1}, but not the input I_t, our model assumes that S_t (a student's knowledge state at a time t) depends on both S_{t-1} (a student's previous knowledge state at $t-1$) and the current input I_t. In other words, we assume that the input (i.e., instructional interventions) impact student knowledge state while KT-IDEM does not.

Beck et al. proposed the HELP model [3] to measure the impact of the tutors' help. The basic structure of the HELP model is very similar to our Intervention-BKT. Note that the input nodes in the Intervention-BKT represents instructional actions (*elicit* vs. *tell*) that are determined by the system. However, in their ITS, help is requested by the student, which may imply the higher knowledge level a student has, the less likely he/she will ask for help. That is, whether the student would ask for help at time t may depend on the student's learned state S_t, which is not reflected in the HELP model. This might be one of the reasons why their results showed that HELP model did not yield a more accurate prediction compared to BKT.

Additionally, a series of research have been done on applying individualized parameters to BKT. For example, Pardos and Heffernan proposed Prior Per Student model[11] which adds a multinomial node representing student's incoming competence to the BKT model and they showed their model performed better. Yudelson et al [13] proposed to use student-specific probability of learning and showed that their method is more effective than BKT. Finally, an innovative approach by Baker and Corbett [2] is to contextually estimate whether each student guesses or slips, thus avoiding the effect of identifiablity and model degeneracy caused by uncertainty. Results showed that their model substantially improved accuracy and reliability compared to the conventional BKT model.

In our paper, we will compare our Intervention-BKT model and BKT against two factor analysis based methods. Previously, BKT has been directly compared against factor analysis based method [8] and the results showed that the latter is as good or better than BKT. However, datasets in their comparisons mainly involve single intervention. In this paper, we will compare all four models on a dataset involving two types of instructional interventions *elicit* and *tell*.

3 Method

3.1 Bayesian Knowledge Tracing (BKT)

BKT is a user modeling method extensively used in ITS. Figure 1 shows a graphical representation of the model and a possible sequence of student observations. The shaded nodes S represent hidden knowledge states. The unshaded nodes O represent observation of students' behaviors. The edges between the nodes represent their conditional dependence.

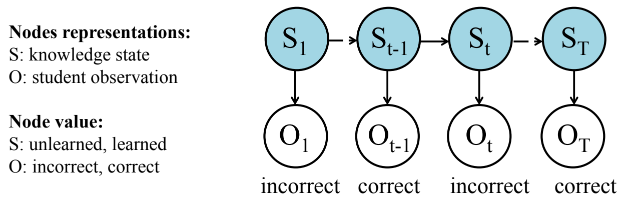

Fig. 1: The Bayesian network topology of the standard Knowledge Tracing model

Fundamentally, the BKT model is a two-state Hidden Markov Model (HMM) [16] characterized by five basic elements: 1) **N**, the number of different types of hidden state; 2) **M**, the number of different types of observation; 3) $\mathbf{\Pi}$, the initial state distribution $P(S_0)$; 4) **T**, the state transition probability $P(S_{t+1}|S_t)$ and 5) **E**, the emission probability $P(O_t|S_t)$. Note that both **N** and **M** are predefined before training occurs, while $\mathbf{\Pi}$, **T** and **E** are learned from the students' observation sequence.

Conventional BKT assumes there are two types of hidden knowledge state (**N**=2), that is, the student's knowledge states (i.e.,*unlearned* and *learned*). It also assumes there are two types of student observation (**M**=2), that is, the student's performance (i.e.,*incorrect* and *correct*). BKT makes two assumptions about its conditional dependence as reflected in the edges in Figure 1. The first assumption BKT makes is a student's knowledge state at a time t is only contingent on her knowledge state at time $t-1$. The second assumption is a student's performance at time t is only dependent on her current knowledge state. These two assumptions are captured by the state transition probability **T** and the emission probability **E**.To fit in the context of student learning, BKT further defines five parameters: 1) **Prior Knowledge** = P(S_0=learned); 2) **Learning rate** = P(learned | unlearned); 3) **Forget** = P(unlearned |learned); 4) **Guess** = P(correct|unlearned) and 5) **Slip** = P(incorrect | learned). Baum-Welch algorithm (or EM method) is used to iteratively update the model's parameters until a maximized probability of observing the training sequence is achieved.

3.2 Intervention Bayesian Knowledge Tracing (Intervention-BKT)

Intervention-BKT is build by incorporating different types of instructional interventions into BKT. Its Bayesian network topology is displayed in Figure 2. Compared with BKT, Intervention-BKT adds a sequence of unshaded input nodes I. The arrows between input nodes I and student observation nodes O represent how instructional interventions affect a student's performance. The arrows between input nodes I and knowledge state nodes S represent how instructional interventions affect a student's hidden knowledge state.

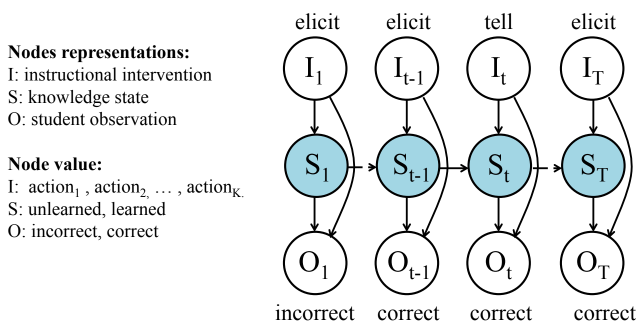

Fig. 2: The Bayesian network topology of the Intervention-BKT model

Intervention-BKT is a special case of Input Output Hidden Markov Model (IOHMM) [17], which is extended from HMM. This model is characterized by six basic elements: 1) **K**,the number of different types of input; 2) **N**,the number of different types of hidden state; 3) **M**,the number of different types of observation; 4) $\mathbf{\Pi}$, the initial state distribution $P(S_0)$; 5) **T**, the state transition probability $P(S_t|I_t, S_{t-1})$ and 6) **E**, the emission probability $P(O_t|I_t, S_t)$

Intervention-BKT makes two distinctions compared to BKT. First, it employs a parameter K representing the number of input types, that is, the instructional intervention types. Second, Intervention-BKT makes two different assumptions about its conditional dependence as represented by the edges in Figure 2: 1) a student's knowledge state at a time t is contingent on her previous state at time $t-1$ **as well as the current intervention** $\mathbf{I_t}$; 2) a student's performance at time t is dependent on her current knowledge state S_t **as well as the current intervention** $\mathbf{I_t}$. Similarly, Our Intervention-BKT employs $1 + 4 \times K$ parameters (compared with 5 parameters of BKT) to describe its conditional probability. The **Prior Knowledge** share the same definition as conventional BKT: **Prior Knowledge**= P(S_0=learned). For each of the K types of interventions $A_j, j \in [1, K]$, Intervention-BKT defines four parameters:

$$\mathbf{Learning\ Rate}_{\mathbf{A_j}} = \text{P(learned|unlearned, } \mathbf{I_t} = \mathbf{A_j}\text{)}$$
$$\mathbf{Forget}_{\mathbf{A_j}} = \text{P(unlearned|learned, } \mathbf{I_t} = \mathbf{A_j}\text{)}$$
$$\mathbf{Guess}_{\mathbf{A_j}} = \text{P(correct|unlearned, } \mathbf{I_t} = \mathbf{A_j}\text{)}$$
$$\mathbf{Slip}_{\mathbf{A_j}} = \text{P(incorrect|learned, } \mathbf{I_t} = \mathbf{A_j}\text{)}$$

In this paper, we mainly focus on modeling two types of instructional intervention *elicit* and *tell*. A possible sequence of instructional interventions is suggested above input node in Figure 2. Note that the conventional BKT model is trained from a sequence of output representing the student's performance, whereas the Intervention-BKT model is trained from a sequence of instructional interventions and the corresponding student's performance.

4 Four Training Corpus

Cordillera [12] is a Natural Language ITS teaching college level introductory physics and all participants in our training corpus experienced identical procedure: 1) completed a survey; 2) read a textbook; 3) took a pretest; 4) solved the same seven training problems on Cordillera, and finally 5) took a post-test. Cordillera provides two types of instructional interventions *elicit* and *tell*. Table 1 demonstrates these two interventions delivering the same domain content.

Table 1: Elicit vs. Tell

(a) **Elicit Version**	(b) **Tell Version**
1. **T:** So let's start with determining the value of v1. 2. **T:** Which principle will help you calculate the rock's instantaneous magnitude of velocity at T1? **{ELICIT}** 3. **S:** definition of kinetic energy	1. **T:** So let's start with determining the value of v1. 2. **T:** To calculate the rock's instantaneous magnitude of velocity at T1, we will apply the definition of kinetic energy again. **{TELL}**

Four training corpus Random, Hybrid, NormGain and InvNormGain were involved in this study. They follow different pedagogical policies with various effectiveness on deciding when to *elicit* and when to *tell*. The remaining components of them are identical. As reported earlier in [15], students learned greatly in all four training corpus, but NormGain students learned in a significantly deeper way, while no difference was found among the rest three. That is to say, the pedagogical policies are most effective in NormGain corpus.

In total, there are 44923 data points from 170 students. More specifically, Random comprises 19584 data points from 64 students; Hybrid comprises 10113 data points from 37 students; NormGain comprises 7691 data points from 37 students and InvNormGain comprises 7535 data points from 32 students. Each student completed around 300 training problem steps. A data point in our training dataset is either the first attempt by students in response to a tutor *elicit*, or a tutor *tells* the next step. The pretest and post-test have the 33 identical test items. All of the tests were graded in a double-blind manner by a single domain expert (not the author). Each test question was assigned two grades: overall and KC-based grade. The overall grade was a score in the range [0, 1] describing the correctness of an answer as a whole, while the KC-based grade was a score in the same range describing the correctness regarding a particular KC.

5 Experiments

Three experiments were conducted. First, we investigated whether Intervention-BKT would outperform BKT on post-test scores prediction. It is commonly considered that relevant knowledge in domains such as math and science is structured as a set of independent but co-occurring Knowledge Components (KCs). A *Knowledge Component (KC)* is "a generalization of everyday terms like concept, principle, fact, or skill, and cognitive science terms like schema, production rule, misconception, or facet" [12]. It is assumed that the student's knowledge state at one KC has no impact on her understanding of any other KCs. This is an idealization, but it has served ITS developers well for many decades as a fundamental assumption made by many student models [7]. In Cordillera, two domain experts identified six primary KCs: Kinetic Energy(KE), Gravitational Potential Energy(GPE), Spring Potential Energy (SPE), Total Mechanical Energy (TME), Conservation of Total Mechanical Energy (CTME), and Change of Total Mechanical Energy (ChTME). We investigated both BKT and Intervention-BKT on each of six primary KCs individually and across KCs (Across). Additionally, to investigate whether the pedagogical strategies would play a role on model performance, we compared them across four datasets individually and combined.

Second, we extended our comparison to two widely applied factor analysis based student modeling approaches: Additive Factor Model (AFM) and Instructional Factor Model (IFM). The original work of [5] has shown IFM outperforms AFM in post-test score prediction. Their experiments were performed by training a KC-general model on Random dataset combined with 10-fold cross-validation. In order for our results to be comparable, we followed the same procedure. The same measurements BIC and 10-fold RMSE were reported.

Third, we explored personalized pedagogical policies suggested by the learned parameters from Intervention-BKT.To make pedagogical recommendations when a student is in the unlearned state, we compared the learning rate (defined as the probability that a student will transit from the unlearned state to the learned state) for elicits and tells (**Learning Rate$_{\mathbf{elicit}}$** vs. **Learning Rate$_{\mathbf{tell}}$**). The tutor action leading to **higher** learning rate will be preferred. Similarly, to make pedagogical recommendations when a student is in the learned state, we compared the forget rate (defined as the probability that a student will transit from the learned state to the unlearned state) for elicits and tells (**Forget**$_{\textbf{elicit}}$ vs. **Forget**$_{\textbf{tell}}$). The tutor action leading to **lower** forget rate is preferred.

6 Results

The forget parameter **Forget** in BKT is fixed to be 0 conventionally, yet prior research showed that BKT may perform better when **Forget** is unfixed [1]. We found using the fixed **Forget** or unfixed **Forget** does not make much difference. Besides, using an unfixed **Forget** can provide us with recommended instructional interventions when students are in the "learned" state. Thus, we will report our models with unfixed **Forget** only.

Three statistics Akaike Information Criterion (AIC), Bayesian Information Criteria (BIC) and the cross-validation Root Mean Squared Error (RMSE) are employed to evaluate our models. For all these measurements, the lower the value, the better the model performs.

6.1 Intervention-BKT vs. BKT

First we will show the results for *Combined* dataset in Table 2. As can be seen, Intervention-BKT produces better model fit than conventional BKT with lower AIC and BIC for all KCs. Furthermore, Intervention-BKT makes more accurate post-test score prediction with LOOCV RMSE at least 0.05 lower than conventional BKT model. For KC ChTME, Intervention-BKT decreases the RMSE by more than 0.2 as marked by "****".

Table 2: Compare Intervention-BKT and conventional BKT on Combined

	AIC		BIC		LOOCV RMSE	
KC	BKT	Intervention-BKT	BKT	Intervention-BKT	BKT	Intervention-BKT
KE	7847	**5343**	7879	**5412**	0.356	**0.252** **
GPE	7712	**5376**	7743	**5445**	0.306	**0.248** *
SPE	3325	**2037**	3356	**2105**	0.419	**0.288** **
TME	7911	**5449**	7943	**5518**	0.347	**0.229** **
CTME	2621	**1620**	2652	**1689**	0.326	**0.254** *
ChTME	2733	**1318**	2764	**1388**	0.471	**0.233** ****
ACROSS	32121	**20495**	32152	**20564**	0.369	**0.275** *

Note: better AIC/BIC in **bold**
*: difference>0.05 ; **: difference>0.1 ; ***: difference>0.15 ; ****: difference>0.2

Next, we compared Intervention-BKT and BKT on four datasets. The same pattern was found in all datasets. Given the space, we only present results for Random in Table 3(a) and NormGain in Table (3(b). Again, both 3 (a) and (b) show Intervention-BKT achieves lower LOOCV RMSE than BKT. For AIC and BIC, Intervention-BKT yields better results for all KCs consistently for Random. It also beats BKT for all KCs for NormGain except for KE and ChTME.

For both Random and NormGain dataset, Intervention-BKT outperforms BKT. However, Intervention-BKT makes greater improvement on Random than

Table 3: Compare Intervention-BKT and conventional BKT on Random and NormGain

	AIC		BIC		LOOCV RMSE	
KC	BKT	Intervention-BKT	BKT	Intervention-BKT	BKT	Intervention-BKT
KE	3687	**2420**	3710	**2467**	0.353	**0.225****
GPE	3026	**1879**	3047	**1926**	0.340	**0.245***
SPE	1637	**953**	1659	**1000**	0.396	**0.267****
TME	3565	**2184**	3586	**2231**	0.326	**0.189****
CTME	1371	**836**	1392	**883**	0.275	**0.211***
ChTME	1160	**607**	1182	**655**	0.453	**0.241*****
Across	14440	**8599**	14462	**8647**	0.373	**0.255****

3(a) Random dataset

	AIC		BIC		LOOCV RMSE	
KC	BKT	Intervention-BKT	BKT	Intervention-BKT	BKT	Intervention-BKT
KE	**1354**	1177	**1371**	1212	0.295	**0.271**
GPE	1407	**1164**	1423	**1199**	0.229	**0.178** *
SPE	491	**397**	507	**432**	0.368	**0.317** *
TME	1413	**1181**	**1429**	1216	0.287	**0.218** *
CTME	346	**257**	362	**292**	0.307	**0.256** *
ChTME	**36**	51	**51**	86	0.523	**0.436** *
Across	5097	**3913**	5113	**3948**	0.342	**0.257** *

3(b) NormGain dataset

on NormGain dataset. More specifically, Intervention-BKT improves BKT by more than 0.1 in five out of seven cases for Random, whereas none of the case shows an improvement greater than 0.1 for NormGain. One possible explanation is Ramdom makes random tutorial decision, whereas NormGain employs pedagogical policies induced by Reinforce Learning (RL). There might be some dependence between pedagogical policies and student's performance, which may cause Intervention-BKT behaves less effective for the NormGain dataset.

6.2 Intervention-BKT vs. BKT vs. AFM vs. IFM

Table 5 shows the performance of the four models Interventio-BKT, BKT, AFM and IFM with respect to post-test score prediction. As we can see, for both BIC and 10-fold RMSE, we have Intervention-BKT > BKT > IFM > AFM and the difference is significant.

6.3 Intervention-BKT parameter analysis

Figure 6 shows the pedagogical suggestion made by Intervention-BKT for all 7 KCs across 5 datasets. Columns 1 to 5 show the suggestions for students who

Table 5: Intervention-BKT vs. BKT vs. AFM vs. IFM on predicting post-test score

Model	BIC	10-fold RMSE
Intervention-BKT	866	0.268
BKT	1170	0.309
IFM	2252	0.453
AFM	2443	0.470

are in the unlearned state. Columns 6 to 10 show the suggestions for students who are in the learned state. As we can see, the recommendations for the system to *tell* or *elicit* vary greatly depending on the KC as well as the training corpus used.

Table 6: Pedagogical suggestion made for KCs when they are unlearned or learned

	Unlearned					Learned				
	R	H	N	I	C	R	H	N	I	C
KE	0.03	0.04	0.083	-0.010	0.021	-0.001	-0.002	-0.043	0.004	-0.010
GPE	-0.002	0.001	0.019	-0.005	0.010	0.001	0.001	0.009	-0.005	-0.004
SPE	0.008	0.003	0.013	0.004	0.006	0.002	0.006	-0.014	-0.005	0.001
TME	-0.009	-0.475	-0.009	-0.031	0.004	0.005	0.000	0.000	0.011	-0.002
CTME	0.005	-0.020	-0.024	-0.002	-0.001	-0.001	0.011	0.011	-0.010	0.001
ChTME	0.022	0.097	0.000	-0.005	0.025	-0.014	0.052	0.000	0.003	-0.013
ACROSS	-0.005	-0.006	-0.007	-0.041	-0.005	0.001	0.000	0.001	0.013	0.001

Tell Elicit

7 Discussion

In this work, we proposed the Intervention-BKT model which incorporates multiple types of instructional interventions into conventional BKT's framework. Our results demonstrated that Intervention-BKT leads to a substantial improvement compared to the BKT model. We also showed that when using RMSE, the former is consistently better than the latter regardless of the effectiveness of the pedagogical policies employed and the difference can be large. Furthermore, we extended our comparison to two other models AMF and IFM and showed Intervention-BKT > BKT > IFM > AFM. Finally, our model showed great potential in closing the loop of instruction design. The learned parameters provide adaptive pedagogical suggestions for students in different learning states. Based on our results, we are confident to say incorporating instructional interventions into BKT enhances model performance significantly, thus it merits further investigation.

For future work, we will investigate the effectiveness of Interventaionl-BKT using other larger datasets from other tutoring systems that may involve multiple instructional interventions, such as *skip* (***elicit*** a question without asking students for explanation) and *justify* (ask students to explain after they give an answer). Additionally, we will experimentally compare the pedagogical policies suggested by Intervention-BKT with our RL-induced policies.

References

1. Beck, J. (2007). Difficulties in inferring student knowledge from observations (and why you should care). In Educational Data Mining: Supplementary Proceedings of the 13th International Conference of Artificial Intelligence in Education (pp. 21-30).
2. Baker, S., Corbett, T., Aleven, V. (2008). More accurate student modeling through contextual estimation of slip and guess probabilities in bayesian knowledge tracing. In ITS (pp. 406-415). Springer Berlin Heidelberg.
3. Beck, E., Chang, M., Mostow, J., Corbett, A. (2008). Does help help? Introducing the Bayesian Evaluation and Assessment methodology. ITS.
4. Cen, H., Koedinger, K., Junker, B. (2006, January). Learning factors analysis–a general method for cognitive model evaluation and improvement. In ITS (pp. 164-175). Springer Berlin Heidelberg.
5. Chi, M., Koedinger, K. R., Gordon, G. J., Jordon, P., VanLahn, K. (2011). Instructional factors analysis: A cognitive model for multiple instructional interventions.
6. Chiappa, S., Bengio, S. (2003). HMM and IOHMM modeling of EEG rhythms for asynchronous BCI systems (No. EPFL-REPORT-82978). IDIAP.
7. Corbett, A. T., Anderson, J. R. (1994). Knowledge tracing: Modeling the acquisition of procedural knowledge.UMUAI, 4(4), 253-278.
8. Gong, Y., Beck, J. E., Heffernan, N. T. (2010). Comparing knowledge tracing and performance factor analysis by using multiple model fitting procedures. In ITS (pp. 35-44). Springer Berlin Heidelberg.
9. Gong, Y., Beck, J. E., Heffernan, N. T. (2010, June). Comparing knowledge tracing and performance factor analysis by using multiple model fitting procedures. In ITS (pp. 35-44). Springer Berlin Heidelberg.
10. Pardos, Z. A., Heffernan, N. T. (2011). BKT-IDEM: Introducing item difficulty to the knowledge tracing model. In User Modeling, Adaption and Personalization (pp. 243-254). Springer Berlin Heidelberg.
11. Pardos, Zachary A., and Neil H. "Modeling individualization in a bayesian networks implementation of knowledge tracing." UMAP. Springer Berlin Heidelberg, 2010.
12. VanLehn, K., Jordan, P. W., Litman, D. (2007). Developing pedagogically effective tutorial dialogue tactics: experiments and a testbed. In SLaTE (pp. 17-20).
13. Yudelson, M. V., Koedinger, K. R., Gordon, G. J. (2013). Individualized bayesian knowledge tracing models. In Artificial Intelligence in Education (pp. 171-180). Springer Berlin Heidelberg.
14. Vanlehn, K. (2006). The behavior of tutoring systems. International journal of artificial intelligence in education, 16(3), 227-265.
15. 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(1-2), 83-113.
16. Eddy, S. R. (1996). Hidden markov models. Current opinion in structural biology, 6(3), 361-365.
17. Marcel, S., Bernier, O., Viallet, J. E., Collobert, D. (2000, March). Hand gesture recognition using input-output hidden markov models. In fg (p. 456). IEEE.