Predicting Students' Performance on Intelligent Tutoring System - Personalized Clustered BKT (PC-BKT) Model

Prema Nedungadi
Amrita CREATE
Email: prema@amrita.edu
Amrita Vishwa Vidyapeetham

M.S.Remya
Dept. of Computer Science and Engineering
Email: ms.remya7@gmail.com
Amrita Vishwa Vidyapeetham

***Abstract*—An Intelligent Tutoring System (ITS) supplements traditional learning methods and is used for personalized learning purposes that range from exploring simple examples to understanding intricate problems. The Bayesian Knowledge Tracing (BKT) model is an established method for student modeling. A recent enhancement to the BKT model is the BKT-PPS (Prior Per Student) which introduces a prior learnt for each student. Although this method demonstrates improved prediction results compared to the others, there are several aspects that limit its usefulness; (a) for a student, the prior learning is common for all skills, however in reality, it varies for each skill (b) Different students have varying learning capabilities; therefore these students cannot be considered as a homogenous group. In this paper, we aim to improve the prediction of student performance using an enhanced BKT model called the PC-BKT (Personalized & Clustered) with individual priors for each student and skill, and dynamic clustering of students based on changing learning ability. We evaluate the predictions in terms of future performance within ASSISTments intelligent tutoring dataset using over 240,000 log data and show that our models increase the accuracy of student prediction in both the general and the cold start problem.**

Keywords—Intelligent Tutoring System(ITS),Bayesian Knowledge Tracing, Personalization, Capability Matrix, Clustering Method.

I. Introduction

Intelligent tutoring systems have been revealed to be highly efficient in helping the students learn better. For example, Shute et al. (1989) claimed that the student who use an ITS for economics required only half time covering the material as compared to a traditional economics course,but perform equally good[6]. The problem of student performance prediction is to predict the likely performance of a student for a given step of a problem which requires specific skills to solve.An ITS can accurately monitor students progress and their learning behavior over time. ITS aims to find the right sequence of items for each learner and uses student models to predict student performance.

Over recent decades, there has been a rich variety of approaches towards modeling student knowledge and skills within interactive learning environments. The Bayesian Knowledge Tracing model [8] is an established method for student modeling. This method can be used to predict the student performance and to find out whether the student has mastered a particular skill.

There exist different variants of BKT including; BKT-EM (Expectation Maximization) [7], BKT-BF (Brute Force) [2], BKT-PPS (Prior Per Student) [14], BKT-CGS (Contextual Guess and Slip) [3]. Although several empirical studies have been conducted to measure which types of student models are better at predicting future performance, both within and outside of the interactive learning environment, the results of the relative performance of different student models has been quite unstable between studies[4].

We reviewed certain models for prediction such as Latent Factor Model [18][19], Multi-Relational Matrix Factorization Model [20], Personalized Forecasting Model [21] and Tensor Factorization model [22][23]. Although these methods demonstrate improved prediction results compared to the others, there are several aspects that limit their usefulness;eg.Different students have varying learning capabilities therefore it is not helpful to consider all students as one group.

In this paper, we aimed to improve the prediction of student performance using a Personalized BKT Model which supports dynamically clustering the students based on changing learning ability. We introduced a factorization model to the BKT model to provide a more accurate performance than the traditional model. The dynamic clustering helps to deal better with the cold-start problem. We evaluated the predictions within ASSISTments intelligent tutoring dataset using over 240,000 log data.

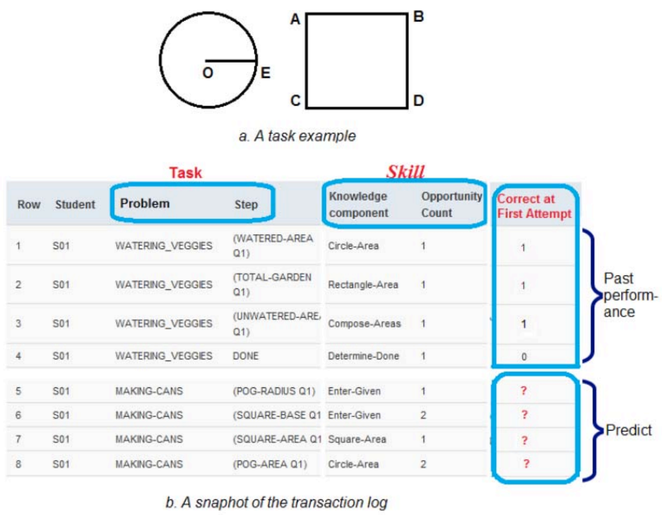
II. Predicting Student Performance in ITS

ITSs are valuable environments for interacting with students in an intelligent way and collecting data for prediction.

A. *Example for Prediction[26]*

Figure 1a presents an example of predicting student performance in an ITS. The task given for the student be, "What is the remaining area of the square after removing an area which is equal to a circular area?" [22] To solve this task (question), students could do some smaller subtasks which are called "steps". Each step may require one or more skills (also called knowledge components), for example:[22]

978-1-4799-3922-0/14/$31.00 ©2014 IEEE



Row	Student	Problem	Step	Knowledge component	Opportunity Count	Correct at First Attempt
1	S01	WATERING_VEGGIES	(WATERED-AREA Q1)	Circle-Area	1	1
2	S01	WATERING_VEGGIES	(TOTAL-GARDEN Q1)	Rectangle-Area	1	1
3	S01	WATERING_VEGGIES	(UNWATERED-ARE/ Q1)	Compose-Areas	1	1
4	S01	WATERING_VEGGIES	DONE	Determine-Done	1	0
5	S01	MAKING-CANS	(POG-RADIUS Q1)	Enter-Given	1	?
6	S01	MAKING-CANS	(SQUARE-BASE Q1	Enter-Given	2	?
7	S01	MAKING-CANS	(SQUARE-AREA Q1	Square-Area	1	?
8	S01	MAKING-CANS	(POG-AREA Q1)	Circle-Area	2	?

Fig. 1: An example of predicting student performance in an ITS [26]

step 1) Calculate the circle area (skill: $area_1 = \pi * (OE)^2$

step 2) Calculate the square area (skill: $area_2 = (AB)^2$

step 3) Calculate the remaining (skill: $area_2 - area_1$).

Each step is recorded as a transaction. Figure 1b presents a snapshot of the transactions[26]. Based on the student's past performance (of the solved tasks), we would like to predict the students' next performance (e.g. correct/incorrect) for some given new tasks.

B. Bayesian Knowledge Tracing Model

We take BKT model as the base model for our work. The Bayesian Knowledge Tracing model [8] is one of the state-of-the-art methods for student modeling. This method can be used to predict performance of the student and to find out whether the student has learned a particular skill.

BKT model has two knowledge parameters, initial knowledge or prior knowledge, denoted as P (L0) and learn rate (transition probability), denoted as P (T), and two performance parameters are guess rate (guess) P (G) and slip rate (slip) P (S). The initial knowledge P (L0) is the probability of Knowledge of a student on a particular skill before interacting with the tutoring system. The learn rate P (T) is the probability that the skill of the student will make the transition from not knowing the skill to having learnt the skill[8]. The guess P (G) is the probability that a student solves the problem correctly even though he/she does not know the skill (the skill is in the unlearned state). The slip rate P (S) is the probability that a student does not solve the problem correctly (making a mistake) even though he/she knows the required skill (the skill is already in the learned state).

Even though the BKT model performs prediction in a reasonably accurate manner, there are two drawbacks. (1) It is not a personalized model and (2) it cannot deal with cold start problems. So we propose a new model which can negate the drawbacks.

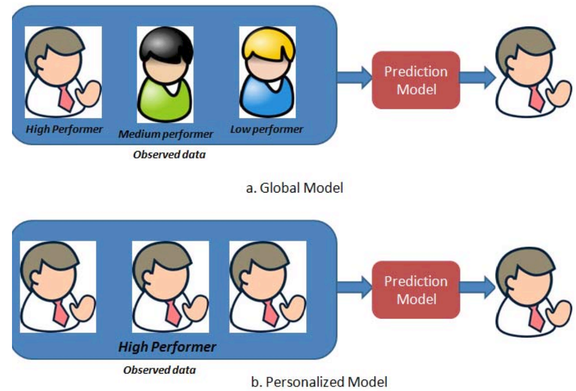


Fig. 2: An example of using personalization

1) Personalized Model [21]: In the Global Model, we build a prediction model on all the observed data and use that model to predict the new data. The proposed method differs from the above approach, in that we only use the historical data of individual students to build the model to predict his/her own performance instead of using all historical data to form the models. Here is an example, including a picture, to describe the reason for our choice:[21].

Using a low performer's results to predict a high performer's outcomes would not fit. We should use either the information of similar students to predict like student's results or we should use an individual's performance to predict his/her outcomes such as in the proposed personalized forecasting methods.

C. Proposed System

The proposed model PC-BKT (EP) is a fully personalized model whereas in the traditional BKT model, the Knowledge Level, P (L) is common for all students. To fit parameters, we use the Empirical Probabilities (EP)[24].

We perform clustering of similar students and skills to deal with the cold start problem. We introduce a capability matrix that has the correct proportion for each skill based on all items attempted by each student involving that skill, expanding on the sum-score work of [11][1].

1) Computing Knowledge Tracing Using Empirical Probabilities(EP)[24]: EP is a two-step process that involves annotating performance data with knowledge, and then using this information to compute the BKT parameters[24].

Annotating Knowledge: The first step in EP is to annotate performance data for each student within each skill with an estimate of when the student learned the skill. We assume there are only two knowledge states: known (1) and unknown (0), and do not allow for forgetting (a known state can never be followed by an unknown state)[24].

Computing the Probabilities :Using the knowledge estimates, we were able to compute each of the BKT parameters, except P(L0), for each skill empirically from the data.

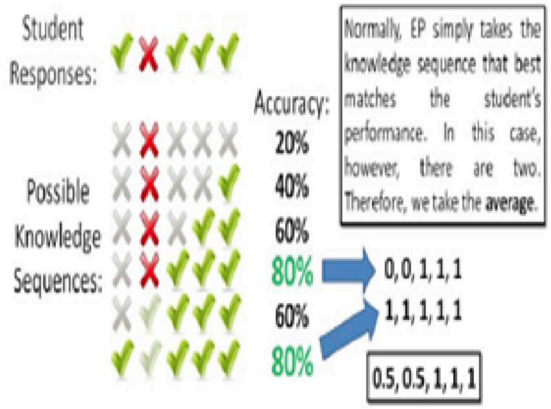


Fig. 3: Each of the six possible knowledge sequences are tried for a students performance history, and in this case, the best two are averaged together to get the final sequence.[24]

2) The Capability Matrix [1]: We begin by assembling the skill dependencies of each item into a Q-matrix [5][17]. The Q-matrix, also referred to as a transfer model or skill coding, is a J $\times$ K matrix where $Q_{jk} = 1$ if item j requires skill k and 0 if it does not, J is the total number of items, and K is the total number of skills[1]. The Q-matrix is usually an expert-elicited assignment matrix.

$$Q = \begin{pmatrix} Q_{1,1} & Q_{1,2} & \dots & Q_{1,K} \\ \vdots & & & \vdots \\ \vdots & & & \vdots \\ Q_{J,1} & Q_{J,2} & \dots & Q_{J,K} \end{pmatrix}$$

$$Y = \begin{pmatrix} Y_{1,1} & Y_{1,2} & \dots & Y_{1,J} \\ \vdots & & & \vdots \\ \vdots & & & \vdots \\ Y_{N,1} & Y_{N,2} & \dots & Y_{N,J} \end{pmatrix}$$

Response matrix Y, N $\times$ J,assembles student response where Y_{ij} indicates both if student i attempted item j and whether or not they answered item j correctly and N is the total number of students[1]. If student i did not attend item j then $Y_{ij} = NA$. The indicator $I_{y_{ij} \neq NA} = 0$ indicates this missing value. If student i attempted item j ($I_{y_{ij} \neq NA} = 1$), then $Y_{ij} = 1$ if they answered correctly, or 0 if they answered incorrectly.

To cluster students by their skill set profiles, we need a summary statistic of their skill performance. We define an N $\times$ K capability matrix B, where B_{ik} is the proportion of correctly answered items involving skill k that student i attempted[1]. If student i did not attempt any items with skill k, we assign a value of 0.5, an uninformative probability of skill mastery[1]. That is, if $\sum_{k=1}^{K} I_{Y_{ij} \neq NA} * Q_{jk} = 0, B_{ik} = 0.5$.Otherwise,

$$B_{ik} = \frac{\sum_{j=1}^{J} I_{Y_{ij} \neq NA} Y_{ij} \cdot Q_{jk}}{\sum_{k=1}^{T} I_{Y_{ij} \neq NA} Q_{jk}}$$

Where Y_{ij} and q_{jk} are the corresponding entries from the response matrix Y and Q-matrix.

There are certain benefits to using a summary statistic of this form. The statistic scales for the number of items in which the skill appears as well as for missing data. If a student has not seen all or any of the items requiring a particular skill, we still derive an estimate based on the available information[1]. Also, the values naturally fall onto a skill mastery scale. For each skill, zero indicates that a student has not learned that skill, one indicates that they have, and 0.5 indicates uncertainty, partial mastery, or no information.

Suppose we have Q-matrix and Y- matrix of training data which contains 3 students and 10 items which can be solved using 2 skills.The the corresponding B-matrix is generated as shown below.

$$Y = \begin{pmatrix} 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & NA & 1 \\ 0 & 0 & NA & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix}$$

$$Q^T = \begin{pmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix}$$

$$\begin{pmatrix} B_{11} = \frac{4}{5} & B_{12} = \frac{1}{5} \\ B_{21} = \frac{1}{5} & B_{22} = \frac{2}{6} = \frac{1}{3} \\ B_{31} = \frac{5}{6} & B_{32} = \frac{6}{6} = 1 \end{pmatrix}$$

D. Algorithm

The algorithm describes the steps involved in our model.

The Model works as follows. The initial Knowledge, (L0), of each student per skill depends mainly on their first performance, because we start with the assumption that the student has maximum knowledge. If the Correct First attempt (CFA) of their first performance is 1, then P (L0) should be 1 minus the probability of guess. Similarly, if it is zero, P (L0) should be equivalent to the probability of slip. The model estimates the guess rate, P(G), the slip rate, P(S) and the learn rate, P (T) of each student by using the Empirical Probabilities[24].

In the training phase, the model updates the Knowledge Level of each student against each skill with respect to the CFA performance. Then, the probability that the student has learned/mastered the skill at opportunity (time) t is computed by

$$P(L_t) = P(L_{t-1}) + (1 - P(L_{t-1})) * P(T)$$

where

$$P(L_{t-1}/x = 1) = \frac{P(L_{t-1}) * (1 - P(S))}{P(L_{t-1}) * (1 - P(S)) + (1 - P(L_{t-1})) * P(G)}$$

$$P(L_{t-1}/x = 0) = \frac{P(L_{t-1}) * P(S)}{P(L_{t-1}) * P(S) + (1 - P(L_{t-1})) * (1 - P(G))}$$

Once the value of the probability reaches one, it will never drop further even if the CFA is Zero. It would affect the performance of the model. So we limit the maximum value of Knowledge Level to 0.95.

The formulas are same as in the BKT, but the difference is that the P(L) is personalized with respect to student against skill. Up to this point, we can use it for predicting the performance of the student. But it doesn't make allowances for the cold start problem. Cold-start problem refers to the issues arising when new students are faced with new tasks involving a skill for which the student has not received training.

Algorithm for Personalised BKT Model

$N = \#students, K = \#skills, J = \#item$

Q- Matrix[item][skill], Y- Matrix[student][item]

Compute Knowledge Sequence(K) Using Empirical Probabilities.

Compute Guess and Slip Using K.

$$P(G) = \frac{\Sigma_j C_j(1-K_j)}{\Sigma_j(1-K_j)}$$

$$P(S) = \frac{\Sigma_j (1-C_j)K_j}{\Sigma_j K_j}$$

Where K_j and C_j are Knowledge and correctness at problem j, respectively.

/*Initial Learning rate of student to each skill*/

for i=1 to N **do**

  for k=1 to K **do**

    if (CFA$_{ik}$ of first performance = 1) **then**

$$P_{ik}(L0) = 1 - P(G)$$

    else

$$P_{ik}(L0) = P(S)$$

    end if

  end for

end for

for i=1 to N **do**

$$P_i(T) = \frac{\Sigma_{j\neq 0}(1-K_{j-1})K_j}{\Sigma_{j\neq 0}(1-K_{j-1})}$$

end for

/*Capability Matrix,B*/

for i=1 to N **do**

  for k=1 to K **do**

$$B_{ik} = \frac{\sum_{j=1}^{J} I_{Y_{ij}\neq NA} Y_{ij} \cdot Q_{jk}}{\sum_{j=1}^{J} I_{Y_{ij}\neq NA} Q_{jk}}$$

Where $I_{Y_{ij}\neq NA}$ is an indicator of elements in Y and can take values 0,1 and NA.

  end for

end for

/*Training Phase*/

for i=0 to N **do**

  for k=0 to K **do**

    for j=0 to J **do**

      if $CFA_t = 1$ **then**

$$P_{ik}(L_{t-1}) = \frac{P_{ik}(L_{t-1}) * (1-P(S))}{P_{ik}(L_{t-1}) * (1-P(S)) + (1-P_{ik}(L_{t-1})) * P(G)}$$

      else

$$P_{ik}(L_{t-1}) = \frac{P_{ik}(L_{t-1}) * P(S)}{P_{ik}(L_{t-1}) * P(S) + (1-P_{ik}(L_{t-1})) * (1-P(G))}$$

      end if

$$P_{ik}(L_t) = P_{ik}(L_{t-1}) + (1-P_{ik}(L_{t-1})) * P_i(T)$$

    end for

  end for

end for

cluster s=KMeansClustering(B)

Calculate Mean P(L) for each cluster.

/*Predicting the performance of Student S on the task required skill i at time t+1*/

$$P_{ik}(L_{t+1}) = P_{ik}(L_t) * (1-P(S)) + (1-P_{ik}(L_{t-1})) * P_(G)$$

TABLE I: KDD CUP 2010 Challenge DataSet

Challenge Data Sets	Students	Steps
Algebra I 2008-2009	3,310	9,426,966
Bridge to Algebra 2008-2009	6,043	20,768,884

In this phase, we perform clustering to deal with the cold start problem. We apply clustering methods to the capability matrix to identify groups of students with similar skill set profiles, similar to [11] which clusters students based on their collaborative behavior[1]. We use the K-Means Clustering algorithm for cluster the students into three clusters, High, Medium and Low. After clustering, we can estimate the average Knowledge Level for each cluster. According to this, we can perform the prediction. The probability that student s will correctly apply a skill at opportunity (time) $t+1$ in a sequence of problem solving is predicted by

$$P(L_{t+1}) = P(L_t) * (1-P(S)) + (1-P(L_{t-1})) * P_(G)$$

III. Testing and Analysis

We want to verify the accuracy of the Personalized BKT Model with clustering using EP (PC-BKT(EP)). For this we compared our model with the traditional BKT Model, the Partially Personalized Model (PP-BKT) ,the Personalized Model without clustering (P-BKT). In the BKT Model, the Knowledge Level is common for all students. In the Partially Personalized Model, the Knowledge Level is personalized but that of a student for all skills is the same. Our model is personalized in all aspects.

Also we compare our model with Personalised BKT Model with clustering (PC-BKT) without using Emperical method(EP).

When a step involves multiple subskills, there are different ways to allocate responsibility among them for observed success or failure of the step. Here we are updating the estimates of skills by giving full responsibility. The ”full responsibility” approach applies these equations to all the subskills.We compare our model with ”update weakest subskill(UWS)[9]” and ”Conjunctive Knowledge Tracing(CKT)[13]” approaches.The update weakest subskill approach simply applies the standard update equations to whichever subskill in a step has the lowest probability, and leaves the others unchanged[25].CKT predicts the probability of a step succeeding as a product of its subskill probabilities[13][25].We modify the CKT as PC-CKT,Personalised CKT with clustering.

Our tests with ASSISTments intelligent tutoring dataset using 200 students log data shows that Personalized BKT Model is significantly more efficient than other models.

A. *Dataset Used for feature selection*

For testing the proposed approached, we use the Cognitive Tutor dataset, from the Knowledge Discovery and Data Mining Challenge 2010 KDD Cup 2010 [12] from two different tutoring systems from multiple schools over multiple school years. The dataset contains 19 attributes. There are mainly 5 data sets: 3 development data sets and 2 challenge data sets from 2 different tutoring systems[12].We used the Challenge dataset for this work. These datasets are quite large which contains 3,310 students with 9,426,966 steps [12] and their original information is described as in Table I & II.There are many technical challenges such as the data matrix is sparse, there is a strong temporal dimension to the data and the problem a given student sees is determined in part by student choices or past success history.

TABLE II: Information of Original Data Sets

Dataset	Size	#Attributes	#Instances
Algebra 2008-2009 train	3.1 GB	23	8,918,054
Algebra 2008-2009 test	124 MB	23	508,912
Bridge to Algebra 2008-2009 train	5.5 GB	21	20,012,498
Bridge to Algebra 2008-2009 test	135 MB	21	756,386

B. *Experimental Setup*

BKT Models were implemented using the C language. The experiment is done with over 240,000 log data from the ASSISTments dataset which contains data logs of 200 students.

C. *Results*

We compare the misclassification percentage of 5 models. First, we test the model with log records containing the same skill in which the students get trained. Figure 4 shows the comparison of 5 models. The graph shows that the % of misclassification is low as compared to the other models. Note that the P-BKT model and PC-BKT Model carries the same percentage. In the absence of cold- start problem, the two models function in the same way. But PC-BKT(EP) Model shows lowest % due to the updating of guess and slip rate.

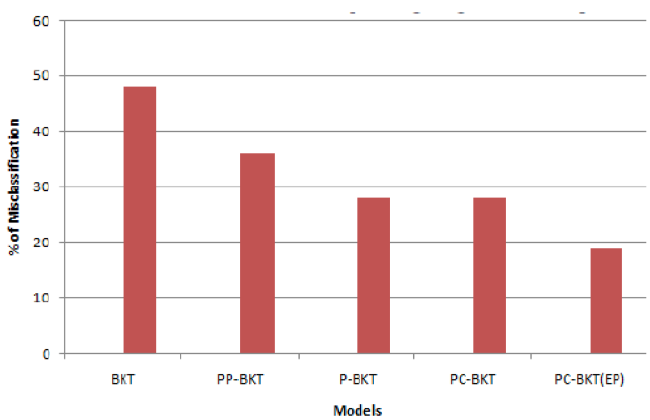


Fig. 4: Comparison of 5 Models with testing data having same skills

Next, we compare the models with testing data involving new skills. Figure 5 shows the comparison of models. The graph reveals that the model with clustering performs better than other models because, only the clustered model has the ability to deal with cold-start problem.

Figure 6 shows the comparison of Misclassification percentage of our model with that of PP-BKT,PC-BKT,PC-CKT and update weakest subskill(UWS) approach.The graph shows that PC-BKT and PC-CKT has almost same accuracy. But our Model PC-BKT(EP) performs better .

IV. Conclusion

We propose three enhancements to the BKT model, the fully personalized P-BKT model ,the PC-BKT, which is the P-BKT model with clustering students into high, medium and low knowledge and the PC-BKT(EP), which is the PC-BKT which uses empirical probabilities to fit the parameters. The P-BKT model has higher student prediction accuracy than both the BKT and the PP-BKT as

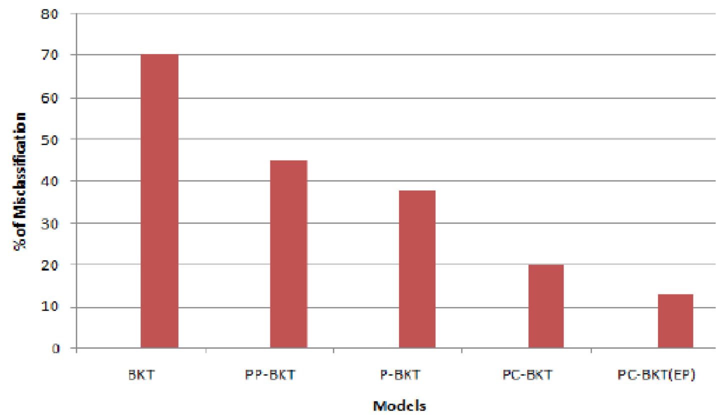


Fig. 5: Comparison of 5 models with cold start problem

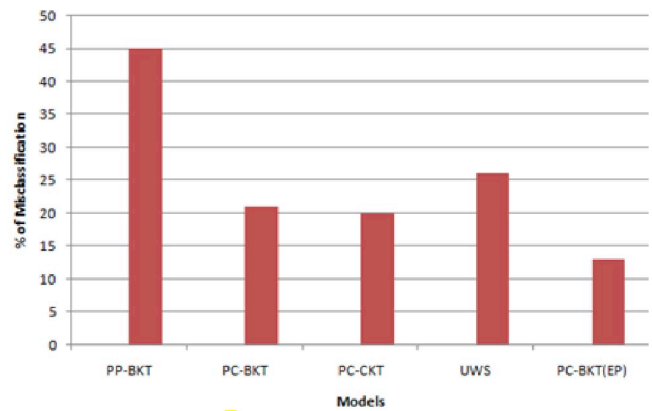


Fig. 6: Comparison of 5 models having different approaches for updating Estimates of skills.

it has individual student priors for each skill. To deal with the cold start problem, of students working on items with skills that have not been encountered yet, we introduce the capability matrix and dynamic clustering based on this and show through extensive data analysis that this model improves the cold start issue. Our tests with ASSISTments intelligent tutoring dataset using over 240,000 log data of 200 students shows that PC-BKT(EP) is significantly more efficient than BKT, P-BKT ,PP-BKT and PC-BKT.

Acknowledgment

This work derives direction and inspiration from the Chancellor of Amrita University, Sri Mata Amritanandamayi Devi. We thank Dr.M Ramachandra Kaimal, head of Computer Science Department, Amrita University for his valuable feedback.

References

[1] E.Ayers, R.Nugent, and N.Dean.(2008) Skill set profile clustering based on student capability vectors computed from online tutoring data. In: EDM2008: 1st International Conference on Educational Data Mining, 20-21 June 2008, Montreal, Canada.

[2] R.S.Baker, A.T.Corbett, and V.Aleven. (2008). More accurate student modeling through contextual estimation of slip and guess probabilities in bayesian knowledge tracing. In Proceedings of the 9th International Conference on Intelligent Tutoring Systems, ITS 08, 406415, Springer-Verlag, Berlin, Heidelberg. 32, 33, 36,37, 38, 40, 50, 51, 54, 65, 78, 81, 92

[3] R.S.J.Baker, A.T.Corbett, S.M.Gowda, A.Z.Wagner, B.A.MacLaren, L.R.Kauffman, A.P.Mitchell, and S.Giguere.(2010). Contextual slip and prediction of student performance after use of an intelligent tutor. In the 18th International Conference on User Modeling, Adaption and Personalization (UMAP 2010), Lecture Notes in Computer Science, 5263, Springer. 35, 36

[4] R.S.J.Baker,Z.Pardos, S.Gowda, B.Nooraei and H.Heffernan.(2011). Ensembling predictions of student knowledge within intelligent tutoring systems. In J. Konstan, R. Conejo, J. Marzo & N. Oliver, eds., User Modeling, Adaption and Personalization, vol. 6787 of Lecture Notes in Computer Science, 1324, Springer Berlin / Heidelberg. 32, 36, 37, 52

[5] T.M.Barnes.(2003). The Q-matrix Method of Fault-tolerant Teaching in Knowledge Assessment and Data Mining. Ph.D. Dissertation, Department of Computer Science, North Carolina State University.

[6] J.Beck, M.Stern, and E.Haugsjaa.(1996). Applications of ai in education. Cross roads, 3, 1115. 1, 2, 3, 13 Chang, K., Beck, J., Mostow, J. and Corbett, A. (2006). A bayes net toolkit for student modeling in intelligent tutoring systems. In Proceedings of International Conference on Intelligent Tutoring Systems (ITS 2006), 104113, Springer. 31, 32, 33, 36, 37, 38, 40, 50, 51, 54, 65, 78, 81, 92

[7] K.Chang, J.Beck, J.Mostow, and A.Corbett.(2006). A bayes net toolkit for student modeling in intelligent tutoring systems. In Proceedings of International Conference on Intelligent Tutoring Systems (ITS 2006), 104113, Springer. 31, 32, 33, 36, 37, 38, 40, 50, 51, 54, 65, 78, 81, 92

[8] A.T.Corbett, and J.R.Anderson. (1995). Knowledge tracing: Modeling the acquisition of procedural knowledge. User Modeling and User-Adapted Interaction, 4,253278. 3, 5, 30, 33, 38, 40, 50, 51, 65, 78, 81, 84, 129

[9] Y.Gong, J. Beck, and N.T. Heffernan.(2010) Comparing Knowledge Tracing and Performance Factor Analysis by Using Multiple Model Fitting Procedures. in Proceedings of the 10th International Conference on Intelligent Tutoring Systems. Pittsburgh, PA: Springer Berlin / Heidelberg.

[10] Jeffrey Dean, and Sanjay Ghemawat.(2004) MapReduce: Simplified Data Processing on Large Clusters. Proceedings of the 6th conference on Symposium on Operating Systems Design and Implementation., pp. 137-150, USENIX Association, Berkley,CA, USA.

[11] J.Henson, R.Templin, and J.Douglas.(2007). Using effcient model based sum-scores for conducting skill diagnoses. Journal of Education Measurement, 44, 361-376.

[12] K.Koedinger, R.Baker, K.Cunningham, A.Skogsholm, B.Leber, and J.Stamper. (2010). A data repository for the edm community: The pslc datashop. In C. Romero, S. Ventura, M. Pechenizkiy and R. Baker, eds., Handbook of Educational Data Mining, Lecture Notes in Computer Science, CRC Press. 5, 13, 16, 18, 20, 26,54, 153

[13] K.R.Koedinger, et al.(2011) Avoiding Problem Selection Thrashing with Conjunctive Knowledge Tracing, in Proceedings of the 4th International Conference on Educational Data Mining.Eindhoven, NL. p. 91-100.

[14] Z.A.Pardos, and N.T.Heffernan.(2010). Using hmms and bagged decision trees to leverage rich features of user and skill from an intelligent tutoring system dataset. In Proceedings of the KDD Cup 2010 Workshop on Improving Cognitive Models with Educational Data Mining, Washington, DC, USA. ix, 25, 34, 36, 37, 40, 153, 154

[15] V.Shute, R.Glaser, and K.Raghaven.(1989). Inference and discovery in an exploratory laboratory. In P. Ackerman, R. Sternberg and R. Glaser, eds., Learning and Individual Differences, 279326, San Francisco: Freeman.

[16] L.Talavera, and E.Gaudioso. (2004). Mining student data to characterize similar behavior groups in unstructured collaboration spaces. Proceedings of the Artificial Intelligence in Computer Supported Collaborative Learning Workshop at the ECAI 2004. Valencia, Spain.

[17] K.K.Tatsuoka.(1983). Rule Space: An Approach for Dealing with Misconceptions Based on Item Response Theory. Journal of Educational Measurement. Vol. 20, No. 4, 345-354.

[18] N.Thai-Nghe, L.Drumond, A.Krohn-Grimberghe, and L.Schmidt-Thieme.(2010). Recommender system for predicting student performance. In Proceedings of the ACM RecSys 2010 Workshop on Recommender Systems for Technology Enhanced Learning (RecSysTEL 2010), vol. 1, 2811 , 2819, Elsevier's Procedia Computer Science.7

[19] N.Thai-Nghe, L.Drumond, T.Horvath, A.Krohn-Grimberghe, A.Nanopoulos, and L.Schmidt-Thieme. (2011). Factorization techniques for predicting student performance. In O.C. Santos and J.G. Boticario, eds., Educational Recommender Systems and Technologies: Practices and Challenges (ERSAT 2011),IGI Global. 7

[20] N.Thai-Nghe, L.Drumond, T.Horvath, and L.Schmidt-Thieme. (2011). Multi-relational factorization models for predicting student performance. In Proceedings of the KDD 2011 Workshop on Knowledge Discovery in Educational Data (KDDinED 2011). Held as part of the 17th ACM SIGKDD Conference on Knowledge Discovery and Data Mining.

[21] N.Thai-Nghe, T.Horvath, and L.Schmidt-Thieme. (2011). Personalized forecasting student performance. In Proceedings of the 11th IEEE International Conference on Advanced Learning Technologies (ICALT 2011), IEEE Computer Society,Athens, GA, USA. 7

[22] N.Thai-Nghe, L.Drumond, T.Horvath, A.Nanopoulos, and L.Schmidt-Thieme. (2011). Matrix and tensor factorization for predicting student performance.In Proceedings of the 3rd International Conference on Computer Supported Education (CSEDU 2011). Best Student Paper Award, 69 78, Noordwijkerhout, the Netherlands. 7, 8

[23] N.Thai-Nghe, T.Horvath, and L.Schmidt-Thieme. (2011). Factorization models for forecasting student performance. In Pechenizkiy, M., Calders, T., Conati, C.,Ventura, S., Romero , C., and Stamper, J. (Eds.). Proceedings of the 4th International Conference on Educational Data Mining (EDM 2011), 11 20, Eindhoven, theNetherlands. 8

[24] W.J.Hawkins, N.T.Heffernan, S.Ryan, and J.D.Baker.(2011) Learning Bayesian Knowledge Tracing Parameters with a Knowledge Heuristic and Empirical Probabilities.Springer-Verlag Berlin Heidelberg.

[25] Y.Xu and J.Mostow. (2012). Comparison of methods to trace multiple subskills: Is LR-DBN best? In Proceedings of the 5th international conference on educational data mining (pp. 4148).

[26] https://pslcdatashop.web.cmu.edu/KDDCup