Dynamic Knowledge Modeling with Heterogeneous Activities for Adaptive Textbooks*

Khushboo Thaker*, Yun Huang*, Peter Brusilovsky, Daqing He
School of Computing and Information
University of Pittsburgh
Pittsburgh, PA, USA
k.thaker,yuh43,peterb,dah44@pitt.edu

ABSTRACT
Adaptive textbooks use student interaction data to infer the current state of student knowledge and recommend most relevant learning materials. A challenge of student modeling for adaptive textbooks is that conventional student models are constructed based on performance data (quiz or problem-solving), however, students' interactions with online textbooks may produce a large volume of student reading data but a limited amount of performance data. In this work, we propose a dynamic student knowledge modeling framework for online adaptive textbooks, which utilizes student reading data combined with few available quiz activities to infer the students' current state of knowledge. The evaluation shows that proposed model learns more accurate students' knowledge state than Knowledge Tracing.

Keywords
student modeling, knowledge tracing, adaptive textbooks

1. INTRODUCTION
Adaptive online textbooks are one of the oldest technologies of personalized web-based learning [7, 10, 16]. A gradual shift to electronic books and textbooks over the last ten years makes this technology even more attractive than in its early days. The challenge for the modern research on adaptive textbooks is its integration with other online learning tools - problems, questions, animations, etc. In particular, student modeling (SM) approaches based on textbook readings behavior should be made compatible with more conventional SM based on student performance. This compatibility would support important "cross-content" recommendation where pages to read could be recommended through the analysis of problem-solving performance while interactive content (animations, problems, questions) could be recommended by considering the reading progress.

*both the authors contributed equally to the paper.(Does NOT produce the permission block, copyright information nor page numbering). For use with ACM_PROC_ARTICLE-SP.CLS. Supported by ACM.

In performance-oriented intelligent tutoring systems (ITS), student knowledge state is measured on the level of individual domain skills or concepts, which are referred to as Knowledge Components (KCs). The main goal of KC-level knowledge modeling is to provide effective learning and reduce the total time of skill acquisition by offering adaptive feedback guiding the student to the most appropriate learning content. To support this personalization, the system keeps track of students' performance such as problem-solving and question-answering. These user interactions are later used by SM systems to distill student knowledge and predict student behavior.

Unfortunately, this well-explored approach could not be directly applied to adaptive textbooks. In most cases, textbook interaction logs provide only a small fraction of performance data (e.g., data on question answering and other activities related to course), which is not sufficient for timely and reliable SM. Naturally, these reading logs provide massive amount of data on student *reading*. However, the use of this data for SM is not straightforward because:

- The reading logs are noisy and not accurate. For example, a student can open a course content, start reading and then switch to some personal task.
- Individual differences (reading proficiency, motivation) could significantly affect student behavior.

In this paper, we present and evaluate a novel approach that combines student activities (reading data and performance data) to construct dynamic student knowledge model for adaptive textbooks. In the remainder of the paper, Section 2 discusses related work; Section 3 describes the proposed approach; Section 4 introduces the evaluation setup; Section 5 presents experimental results; and Section 6 summarizes conclusions and directions of future work.

2. RELATED WORK
2.1 Knowledge Tracing in ITS
Knowledge Tracing (KT) model was introduced in 1995 by Corbett and Anderson [3]. KT uses Hidden Markov Models (HMM) to represent student knowledge as binary latent variables. Each latent variable represents student knowledge of a particular KC, which could be either known or unknown. The observed variable is the performance of student at a given step, which is measured as a binary variable representing the correctness of a step or an answer (correct or not correct). KT directly represents KC-level knowledge estimation and allows dynamic knowledge update at each student learning opportunity. The conventional KT model

has been extended further to learning individualized features [13] and providing instructional based intervention node [12]. In this work, we follow the KT modeling approach since we need knowledge estimates of different KCs to support several kinds of personalization.

2.2 Adaptive Online Textbooks

The research on adaptive textbooks has been motivated by the increasing popularity of World Wide Web (WWW) and the opportunity to use this platform for learning. The hypertext nature of early WWW made an online hypertext-based textbook a natural media for learning while the increased diversity of Web users stressed the need for adaptation. The first generation of adaptive textbooks [2, 4, 7, 10] focused on tracing student reading behavior to guide students to most relevant pages using adaptive navigation support [2, 4, 7, 16] or recommendation [10]. These types of personalization were based on a sophisticated knowledge modeling: each textbook page was associated with a set of concepts *presented* on the page as well as concepts *required* to understand the page [2, 4]. On the other hand, SM was relatively simple: these systems treated each visit to a page as a contribution to learning all presented concepts.

A significant trend of modern online textbooks is the increased inclusion of interactive content "beyond text". While the attempts to integrate online reading with problem solving have been made in the early days of online textbooks [16], it was a rare exception. Modern textbooks, however, routinely integrate a variety of "smart content" such as visualizations, problems, and videos. In this context, the ability to integrate data about student work with all these components and use it for a better-quality SM becomes a challenge for modern online textbooks.

3. KNOWLEDGE MODELING IN ADAPTIVE TEXTBOOKS

Our work attempts to combine the ideas of reading-based SM explored in the area of adaptive textbooks with the ideas of performance-based modeling explored by conventional ITS. The goal is to develop more reliable modeling for modern adaptive textbooks that could support several kinds of personalization such as guiding students to most appropriate sections or recommending relevant external content. This section introduces our earlier work on SM in textbooks and presents two novel models that combine reading-based KT [9] with performance-based KT [3] thus leveraging both reading and question-answering data.

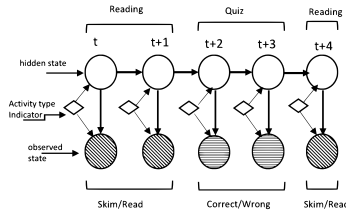
3.1 Behavior Model (BM) and Its Problems

As a baseline model in this work we use, Behavior Model (BM) suggested and explored earlier by Huang et al. [9]. The BM has a strict assumption that students reading speed is positively correlated with their knowledge state. However, other research indicated that this assumption might not always hold [1]. Indeed, in the dataset we considered for this study we observed a negative correlation between student reading behavior and quiz performance of -0.58, which indicates that data consists of mixture different types of students with noisy reading interactions. The primary goal of models presented in this paper was to improve BM. Our key ideas are (1) to handle mixture and noisy reading behavior among students by tuning it with other available activities performed by the student and (2) incorporate individual student differences to address better knowledge estimation for different types of students. In two following subsections, we present two models that advance the original BM in the proposed directions.

3.2 Behavior-Performance Model (BPM)

To achieve this we utilized Feature Aware Student Knowledge Tracing (FAST) framework [11], which replaces the conditional probability tables of the emission and transmission probabilities in BM framework with logistic regression (LR) distribution. HMM parameters are thus computed based on LR with features at each time step. This allows flexibility of incorporating a large number of features at each learning step. To enable FAST for different types of observation variables we introduce an activity type indicator variable which is set to 0 for *Read* and 1 for *Skim* (see Figure 1).

Figure 1: Behavior Performance Model (BPM)



3.3 Individualized Behavior-Performance Model (IBPM)

The *BPM* incorporates reading activities as binary variables with values *Skim* and *Read*. Since reading is a continuous variable, discretization of this manner causes a lot of information loss at student level. This information might be very helpful to characterize individualized student reading behavior and to obtain individualized parameters for different kinds of students. We propose *Individualized Behavior-Performance Model (IBPM)* that incorporates the individualized reading speed information as a feature in addition to activity type indicator features. This feature is based on accumulated median reading speed from first reading activity till $(t-1)$th reading activity of a student, where t is the current step of observation in an HMM of a KC. The feature is normalized to be in the range of 0 to 1 as there is a large variance in reading speed observation. Thus at each step along with different activity sequence observed, the model is also provided individual average reading speed observed so far. There are several benefits of our method:

- This method provides different sets of parameters (learn, guess, slip) for students with different reading speed.
- Compared with adding a parameter per-student for individualization, this feature provides more generalized modeling, because it learns the in-general association of the speed with HMM parameters for each KC.
- It is a flexible approach to integrate other behavior features as FAST has linear complexity in respect to the number of features [11].

4. EXPERIMENTS

4.1 System and Dataset

The dataset used for the experiment is collected from online reading platform Reading Circle [6] in spring 2016. This system was used for a graduate level course on Information Retrieval at the University of Pittsburgh. The system provides an active reading environment where students read the material of the assigned textbook to prepare for the next class. Each section of the assigned reading is followed by a quiz with several questions, which allow students to assess how well they learned the content. There is no restriction on the number of attempts to the questions. The final dataset contains 22,536 interactions from 22 students (see Table 1).

Table 1: Dataset Statistics

documents	394
questions	158
Average questions attempted	126
% of skimming Activities	33
% of reading Activities	67

4.2 Data-Preprocessing

Discretization of reading time is performed to label the observations to *Read* and *Skim*. For discretization we followed the same technique as performed by Huang et al. [9] The key to well-trained KT model is to have correct representative KCs. The conventional way of defining KCs is manual knowledge modeling by subject experts. Recently, Huang et al. [9], tried different KC extraction methods and found automatic word-based method to be reliable. However, word-based method gives a large set of KCs and it is very noisy. To improve automatic KC extraction based on words' importance in a reading unit, we applied the TF*IDF (Term Frequency - Inverse Document Frequency) approach. For each document, top 5 TF*IDF-weighed words were extracted and considered as KCs for that reading.

4.3 Tools and Parameters

For building both BPM and IBPM models, we used open source FAST toolkit [5]. HMM models are prone to get trained for local optimum values, due to which proper initialization of HMM parameters is very important. In all the models the HMM modes were initialized with (0.1,0.1,0.8,0.8) parameter values for ($P(L_0)$,$P(T)$,$P(G)$,$P(S)$). This choice of initialization is based on observing the negative correlation between reading and performance and preliminary experiments under another initial parameter set (0.1,0.1,0.2,0.2) where the predictive performance of all models was worse [9].

4.4 Baseline Methods

In order to show the performance gain of proposed approach, we used two variations of KT as baselines. The first model is the *Behavior Model* (**BM**) reviewed in section 3.1, and the second is *Performance Model* (**PM**) trained on quiz activities by the student. In addition we use a majority class baseline (**MC**). As the proposed model is able to perform both reading time and quiz performance predictions, **BM** and **PM** separately act as a baseline for proposed models' reading time prediction and quiz performance prediction task.

4.5 Cross Validated Prediction Evaluation

FAST trains individual HMM for each KC using training data and performs prediction on test data. Firstly, we randomly selected 50% of students and put all their reading and quiz activity data into training set. Then for the remaining 50% of students, we put the first half of their activity sequence into training set. The second half of their activity sequences are withheld for test set. This process is repeated 10 times. The prediction is reported on reading speed, first attempt quiz performance, and all-attempts quiz performance. 10 split cross-validation is performed from the generated folds. Both Area Under the Receiver Operating Characteristic curve (AUC) and Root Mean Squared Error (RMSE) are reported based on a recent paper, that raised a concern about using only AUC for evaluation of SM [14].

5. RESULTS AND DISCUSSION

5.1 Predictive Performance of BPM

Table 2 summarizes the predictive performance computed by averaging across 10 splits and Table 3 reports significance. Comparing with *MC*, *BPM* has significantly better RMSE and AUC across all prediction tasks. The relatively lower AUC value of *BPM* in reading prediction task indicates high noise in reading interactions. Since quiz performance usually correlates better with knowledge than reading behavior, the prediction on quiz is of more importance than that on reading, thus the result indicates a clear advantage of *BPM* over *MC*. Comparing with *BM* and *PM* which are trained on a single type of interactions, *BPM* also beats them significantly in corresponding prediction tasks in both RMSE and AUC metrics. We clearly see the advantage of integrating behavior and performance data in *BPM* over *PM* and *BM*. Better performance of *BPM* over *BM* indicates that even a small amount of quiz performance data could significantly improve knowledge inference and performance prediction. Better performance of *BPM* over *PM* indicates that reading data albeit being noisy still carries valuable information that could help infer knowledge and conduct prediction.

5.2 Predictive Performance of IBPM

The intuition behind *IBPM* is that it provides additional student reading behavior features (in addition to activity type indicator) for capturing individual differences. As can be seen in Table 2, *IBPM* incorporating individualized speed feature shows improvement by both RMSE and AUC metrics compared with *BPM*. The improvement is significant for reading speed prediction task and quiz all-attempts performance prediction. However, its improvement over *BPM* on predicting first attempt performance in terms of RMSE is not significant. A probable reason is that our dataset exhibits a mixture of students in terms of reading behavior and performance (indicated by negative correlation value).

Table 2: Prediction performance for reading speed, 1st attempt quiz prediction, and all attempts. Two best results are shown in bold.

Model	RMSE	AUC	RMSE	AUC	RMSE	AUC
	reading		1st att.		all att.	
IBPM	**.483**±.008	**.512**±.014	**.472**±.004	**.635**±.018	**.391**±.007	**.867**±.010
BPM	**.487**±.008	.458±.012	**.473**±.004	**.633**±.018	**.391**±.007	**.867**±.010
BM	.508±.011	.442±.019	-	-	-	-
PM	-	-	.504±.002	.602±.014	.427±.005	.803±.009
MC	.593±.019	**.500**±.000	.550±.013	.500±.000	.693±.003	.500±.000

Table 3: Paired t-test p value for reading and quiz prediction performance with Bonferroni correction

	read		1st att.		all att.	
Compared Models	RMSE	AUC	RMSE	AUC	RMSE	AUC
IBPM vs *BPM*	***	***	0.18	*	*	*
IBPM vs *BM/PM*	***	***	***	***	***	***
IBPM vs *MC*	***	**	***	***	***	***
BPM vs *BM/PM*	***	***	*	***	***	*
BPM vs *MC*	***	***	***	***	***	***

10CV paired t-test, p-values
$^{*}0.05/5 = 0.01$, $^{**}0.01/5 = 0.002$, $^{***}0.001/5 = 0.0002$

5.3 Parameter Analysis of BPM

To validate our hypothesis that quiz activities contain less noise than reading activities for inferring knowledge, we conduct a drill-down analysis of parameters of *BPM* and baseline models. We compute the parameters for each KC in

BPM by setting the value of activity type indicator to 0 for the reading part and 1 for quiz part in the logistic regression of each parameter, and then average the parameters across all KCs. According to Table 4, BPM has fitted lower *guess* and *slip* parameters in quiz activity part than reading activity part, which indicates that quiz activities have higher positive correlation with knowledge state than reading activities i.e., quiz activities indeed have much less noise for inferring knowledge. In addition, Table 4 shows that the parameters learned for guess and slip for BPM are smaller than those for BM and PM, which indicates that BPM has higher plausibility enabling more accurate knowledge inference than these baseline models [8]. The high values of *guess* and *slip* parameters for BM and PM model indicates that single activity is not able to learn accurate student behavior.

Table 4: Parameters learned by different models for learn, guess and slip probabilities

Model	Activity Type	learn	guess	slip
BM	Reading	0.384	0.505	0.776
PM	Quiz	0.091	0.705	0.589
BPM	Reading	0.404	0.363	0.420
BPM	Quiz	0.354	0.288	0.313

6. CONCLUSION AND FUTURE WORK

This paper investigated the significance of integrating heterogeneous student activities in a KT framework for adaptive textbooks. The integrated model BPM was trained with large volume of noisy reading data and small amount of quiz performance data. BPM significantly outperforms the basic model BM, which is based on only reading behavior logs, and PM which is based on only quiz behavior logs. The results indicate that combining quiz and reading interactions help in inferring student knowledge state. To address student differences, $IBPM$ integrated continuous observation in BPM. The performance of $IBPM$ was similar to BPM with a considerable improvement on reading speed prediction and small improvement on quiz performance prediction. In the future, we would like to further investigate $IBPM$ by utilizing other individualization features.

Although overall performance is not as high as in ITS focused on mastery learning, our past experience with topic-based SM [15] hints that current level of prediction performance could be sufficient to deliver successful personalization based on adaptive navigation support where the student can choose from several recommended options. We plan to assess the value of our SM approach as a basis for personalized guidance in the future studies.

Our work could be considered as the first attempt to model dynamic student knowledge in adaptive textbooks with heterogeneous interactions. We believe that the possibility of integrating individual differences to the proposed model makes it especially promising for real-time learning systems. Moreover, our approach makes it possible to integrate more types of student activities like search, video, listening and discussion to further increase the quality of modeling and to provide holistic SM. We plan to explore these opportunities in the future work.

7. REFERENCES

[1] R. S. Baker, A. T. Corbett, K. R. Koedinger, and A. Z. Wagner. Off-task behavior in the cognitive tutor classroom: When students "game the system". In *Proc. ACM Conf. on Human Factors in Computing Systems*, CHI '04, pages 383–390, 2004.

[2] P. Brusilovsky and J. Eklund. A study of user-model based link annotation in educational hypermedia. *J. of Universal Computer Science*, 4(4):429–448, 1998.

[3] A. T. Corbett and J. R. Anderson. Knowledge tracing: Modelling the acquisition of procedural knowledge. *User Modeling and User-Adapted Interaction*, 4(4):253–278, 1995.

[4] P. De Bra. Teaching through adaptive hypertext on the www. *Int. Journal of Educational Telecommunications*, 3(2/3):163–180, 1997.

[5] J. González, Y. Huang, and P. Brusilovsky. General features in knowledge tracing: Applications to multiple subskills, temporal item response theory, and expert knowledge. In *Proc. the 7th Int. Conf. on Educational Data Mining*, pages 84–91, 2014.

[6] J. Guerra, D. Parra, and P. Brusilovsky. Encouraging online student reading with social visualization. In *The 2nd Workshop on Intelligent Support for Learning in Groups at the 16th Conf. on Artificial Intelligence in Education*, pages 47–50, 2013.

[7] N. Henze, K. Naceur, W. Nejdl, and M. Wolpers. Adaptive hyperbooks for constructivist teaching. *Künstliche Intelligenz*, 13(4):26–31, 1999.

[8] Y. Huang, J. González, R. Kumar, and P. Brusilovsky. A framework for multifaceted evaluation of student models. In *Proc. the 8th Int. Conf. on Educational Data Mining*, pages 203–210, 2015.

[9] Y. Huang, M. Yudelson, S. Han, D. He, and P. Brusilovsky. A framework for dynamic knowledge modeling in textbook-based learning. In *Proc. 24th Conf. on User Modeling, Adaptation and Personalization*, pages 141–150, 2016.

[10] A. Kavcic. Fuzzy User Modeling for Adaptation in Educational Hypermedia. *IEEE Transactions on Systems, Man and Cybernetics*, 34(4):439–449, 2004.

[11] M. Khajah, Y. Huang, J. González, M. C. Mozer, and P. Brusilovsky. Integrating knowledge tracing and item response theory: A tale of two frameworks. In *Workshop on Personalization Approaches in Learning Environments at Int. Conf. on User Modeling, Adaptation, and Personalization*, pages 7–12, 2014.

[12] C. Lin and M. Chi. Intervention-BKT: Incorporating instructional interventions into bayesian knowledge tracing. In *Intelligent Tutoring Systems*, pages 208–218, Cham, 2016. Springer.

[13] Z. A. Pardos and N. T. Heffernan. Modeling individualization in a bayesian networks implementation of knowledge tracing. In *Proc. the 18th Int. Conf. on User Modeling, Adaptation, and Personalization*, pages 255–266. Springer-Verlag, 2010.

[14] R. Pelánek. Metrics for evaluation of student models. *Journal of Educational Data Mining*, 7(2):1–19, 2015.

[15] S. Sosnovsky and P. Brusilovsky. Evaluation of topic-based adaptation and student modeling in quizguide. *User Modeling and User-Adapted Interaction*, 25(4):371–424, 2015.

[16] G. Weber and P. Brusilovsky. ELM-ART: An adaptive versatile system for web-based instruction. *Int. Journal of Artificial Intelligence in Education*, 12(4):351–384, 2001.