Structured Knowledge Tracing Models for Student Assessment on Coursera

Zhuo Wang
School of EECS
Peking University
Beijing, 100871, China
wangzhuo@pku.edu.cn

Zhiting Hu
School of Computer Science
Carnegie Mellon University
Pittsburgh, PA 15213
zhitingh@cs.cmu.edu

Jile Zhu
School of EECS
Peking University
Beijing, 100871, China
zhujile0918@pku.edu.cn

Ming Zhang
School of EECS
Peking University
Beijing, 100871, China
mzhang@net.pku.edu.cn

Xiang Li
School of EECS
Peking University
Beijing, 100871, China
lixiang.eecs@pku.edu.cn

L@S 2016, April 25–26, 2016, Edinburgh, UK.
ACM 978-1-4503-3726-7/16/04.
http://dx.doi.org/10.1145/2876034.2893416

Abstract
Massive Open Online Courses (MOOCs) provide an effective learning platform with various high-quality educational materials accessible to learners from all over the world. However, current MOOCs lack personalized learning guidance and intelligent assessment for individuals. Though a few recent attempts have been made to trace students' knowledge states by adapting the popular Bayesian Knowledge Tracing (BKT) model, they have largely ignored the rich structures and correlations among knowledge components (KCs) within a course. This paper proposes to model both the hierarchical and the temporal properties of the knowledge states in order to improve the modeling accuracy. Based on the content organization characteristics on the Coursera MOOC platform, we provide a well-defined KC model, and develop Multi-Grained-BKT and Historical-BKT to capture the above features effectively. Experiments on a Coursera course dataset show our approach significantly improves over previous vanilla BKT models on predicting students' quiz performance.

Author Keywords
Knowledge Tracing; MOOCs; Student Assessment; Student Modeling; Hierarchical and Temporal

Parameter	Explanation
p(L$_0$)	Learned: initial knowledge states
p(T)	Transition: the probability of learning the knowledge component
p(G)	Guess: the probability of guessing correctly in the unlearned state
p(S)	Slip: the probability of answering incorrectly in the learned state

Table 1: All the four parameters in basic BKT.

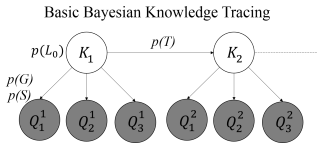

Figure 1: Basic BKT: a time slice represents a quiz submission and there are multiple questions in a single time slice. Node K_t represents the student's knowledge state at time t and node Q_k^t represents the student's response for question k at time t.

Introduction

MOOCs as an effective learning platform provide abundant high-quality learning resources. With the increasing prevalence, however, it becomes impossible for instructors to track individual learners' knowledge states and provide personalized learning guidance. An automatic and accurate student assessment model can be useful for both instructors and learners. For example, it can provide overall feedback to instructors and suggest the need for further explanations; and infer the weakness of students and recommend appropriate learning materials.

Extensive advances have been made on student assessment. The BKT model [1] is among the most popular techniques, especially in the intelligent tutoring systems where knowledge is manually categorized into fine-grained KCs by experts. Despite its impressive success, application of BKT on MOOCs is limited due to the lack of explicit KC definition. A few recent work has attempted to retrofit BKT: e.g., Pardos et al. adapted BKT to edX [1] by assuming questions in quizzes as KCs [3]. However, these simple extensions largely fail to model the rich structures and correlations among KCs within a MOOC, and thus can lead to suboptimal performance. This paper addresses the above issues on Coursera [2], one of the most popular MOOC platforms. We leverage the resource organization characteristics where each chapter of a course consists of multiple lecture videos, based on which we provide a well-defined KC model, and further develop two novel knowledge tracing methods, Multi-Grained-BKT and Historical-BKT, to model the distinct yet closely-related fine-grained KCs for each coarse-grained KC, and capture the correlations between multiple quiz submissions, respectively. Both methods significantly outperform previous vanilla BKT models on a dataset collected from a Coursera course. To the best of our knowledge, we are the first to model both the hierarchical (i.e., multi-grained) and the temporal (i.e., historical) properties of knowledge states for accurate student assessment. Our method can be helpful for integrating adaptive learning systems and designing personalized study plans on MOOCs.

[1] https://www.edx.org/
[2] https://www.coursera.org/

Methodology

Basic Bayesian Knowledge Tracing Model

Knowledge tracing was proposed based on mastery learning where knowledge is composed of KCs. The notion was then introduced into intelligent tutoring [1]. BKT is a hidden Markov model (HMM), as shown in Figure 1, where student's knowledge states are represented by a series of binary variables K. There are four key parameters in BKT (Table 1), namely, prior knowledge, probability of learning, guessing, and slipping. And we have:

$$p(Q_k^t = 1) = p(L_t)(1 - p(S)) + (1 - p(L_t))p(G)$$
$$p(L_t = 0) = p(L_{t-1}) + (1 - p(L_t))p(T)$$

The parameters are learnt from data through the expectation-maximization (EM) algorithm. BKT is widely used in the context of intelligent tutoring systems, where KCs have been designed in detail by domain experts, and each question is labelled with a specific KC.

Knowledge Organization on Coursera

In order to trace knowledge states on MOOCs using BKT, we first notice that Coursera (and most other MOOC platforms) allows multiple submissions for a quiz and we regard each submission as a time slice. Moreover, Coursera allows several variations for a question so that questions are different between sequential trials for the same quiz. In fact, students are encouraged to do quizzes again and again to master relevant knowledge.

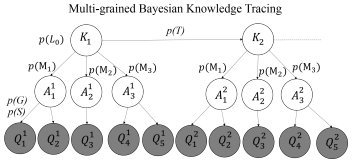

Figure 2: Multi-Grained-BKT: there are multiple fine-grained KCs for a coarse-grained KC with one or more questions. Node A_k^t represents the knowledge state of fine-grained KC k at time t. $p(M_k)$ represents the probability of mastering fine-grained KC k when the overall KC has been mastered.

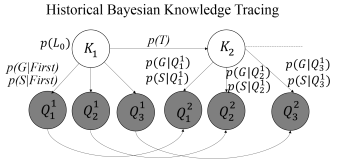

Figure 3: Historical-BKT: The response depends not only on the KC, but also on the previous question response. That is, $p(G)$ and $p(S)$ parameters of node Q_k^t depend on node Q_k^{t-1}

Compared to the mastering learning context, on MOOCs there is no explicit KC definition. A straightforward solution is to regard a whole chapter as a KC, because each chapter is about a specific topic. We call the baseline BKT model based on this definition as *Coarse-BKT*. However, such a KC definition is too coarse and it is hard to model student's knowledge states precisely.

An alternative way to enable fine-grained tracing is to make use of the characteristics of the knowledge organization on MOOCs where each chapter usually consists of several lecture videos and each video, often in a few minutes, focuses on a specific yet complete topic. We can thus see the content covered by each video as a KC. We denote the second baseline, which applies the Basic BKT on the fine-grained KC definition, as *Fine-BKT*.

Such a KC definition has several advantages: 1) MOOCs usually provide abundant materials for lecture videos, e.g., slides and in-video quizzes. This makes KCs well-defined and interpretable, as compared to clustering-based methods [2]; 2) Computers can handle KCs through natural language processing techniques automatically using texts like subtitles and forum posts. This paper asked two teaching assistants who had designed the quizzes to do the labelling for accuracy.

Multi-Grained-BKT

However, Fine-BKT ignores the fact that KCs in a chapter are closely related to each other. To address this issue, we come up with a novel method called *Multi-Grained-BKT*, in which we still regard the whole chapter as an overall coarse-grained KC and the content covered by each lecture video as one of its fine-grained KC.

Figure 2 shows Multi-Grained-BKT has an additional layer of "fine-grained KCs". When a student masters a coarse-grained KC, there is a probability of mastering one of its fine-grained KCs, modelled by $p(M_k)$. Otherwise, he/she cannot master any of its fine-grained KCs, i.e.,

$$p(A_k^t = 0) = p(L_t)(1 - p(M_k)) + (1 - p(L_t))$$

The parameter $p(M_k)$ models the difficulty of mastering a specific fine-grained KC. We pre-specify its value as the average correct rate of relevant questions:

$$p(M_k) = \frac{\#\{correct\ responses\ for\ fine-grained\ KC\ k\}}{\#\{all\ responses\ for\ fine-grained\ KC\ k\}}$$

We thus control the model complexity to four free parameters as before, and use EM algorithm to estimate the parameters and predict quiz submission results.

Historical-BKT

As mentioned, each submission is regarded as a time slice, and Coursera allows variations for a question in sequential trials. However, these questions are still closely related, usually about the same problem with different operational data or different multiple choice options. To make use of this characteristic, we propose a novel method called *Historical-BKT* (Figure 3). The probability of guessing and slipping will depend on the previous question response. Intuitively, if previous response is correct $p(G)$ tends to be larger and $p(S)$ smaller. We thus have:

$$p(Q_k^t = 1) = p(L_t)(1 - p(S|Q_k^{t-1})) + (1 - p(L_t))p(G|Q_k^{t-1})$$

As there are three different states for responses in the previous submission (i.e., Correct, Incorrect, and Not existing), this model has eight parameters in total.

Experiment

Setup

Our data is from the Data Structures and Algorithms course held by Peking Univ. on Coursera in fall, 2013. It

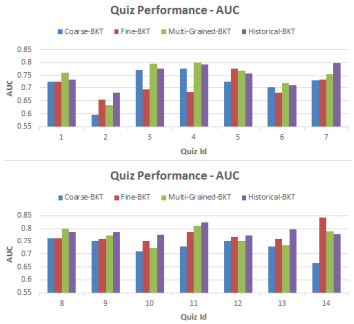

Figure 4: Cross-validated AUC results of the four models on all 14 quizzes. Multi-Grained-BKT and Historical-BKT achieve higher AUC in most quizzes.

Model	Coarse	Fine	Multi-Grained	Historical
Average AUC	0.7233	0.7415	0.7584	0.7695
P-value to Coarse-BKT	—	0.3110	0.0015	0.0003

Table 2: The average AUC results on all 14 quizzes and the p-values of t-tests between Coarse-BKT and the other three.

	Previous Wrong	Previous Correct	First Submission
p(G)	0.3099	0.5687	0.5178
p(S)	0.3211	0.0578	0.1683

Figure 5: Parameter analysis for Historical-BKT shows that $p(G|PreviousCorrect) > p(G|PreviousWrong)$ and $p(S|PreviousCorrect) < p(S|PreviousWrong)$

.

was a 14 week-long online course with learning materials published in each week as a chapter. Weekly quizzes consist of an average of 7.43 questions. There are 13,170 students registered for this course. We selected those who had submitted quizzes, and got a dataset of 1,077 students and 6,583 submissions. The KC labels of quiz questions were acquired from the teacher assistants.

We compare all four models on predicting students' question responses in the last submission with a 5-fold cross-validation at student level. [3] We use the area under the curve (AUC) as the metric, which is widely-used in binary classification. We enumerate all pairings of correct and incorrect responses and the AUC is simply the percentage of the pairings where the correct responses get higher predictive probability over the incorrect responses.

Result

Figure 4 shows the AUC results of the four models on all the 14 quizzes, and Table 2 lists the two-tailed paired t-tests to evaluate the significance of the performance difference. The AUC of Fine-BKT is not significantly different from that of Coarse-BKT, showing that merely using the fine-grained KC is not sufficient. In contrast, Multi-Grained-BKT achieves an improvement of 0.0351 over Coarse-BKT (p-value=0.0015). The reason is that Multi-Grained-BKT models the hierarchical structure of KCs in different granularities, and is able to capture the relatedness between fine-grained KCs in the same chapter.

Historical-BKT, which captures the relations between multiple quiz submissions, performs as well as the Multi-Grained-BKT (p-value=0.1801) and achieves a significant improvement of 0.0462 over Coarse-BKT (p-value=0.0003). Figure 5 shows that students are more likely to response correctly if their answer was already correct in the previous submission.

Conclusion and Future Work

This paper has provided a new KC definition on MOOCs and developed two knowledge tracing methods, Multi-Grained-BKT and Historical-BKT, to model the hierarchical KCs and capture the correlations between quiz submissions, respectively. Both methods greatly improved over the vanilla BKT. We note that our methods rely on those features common to popular MOOC platforms (e.g., Coursera and edX), e.g., allowing multiple quiz submissions with slight question variations. This ensures the applicability of our methods on these platforms.

There are several directions worth further investigation in the future. A complicated question in a quiz can be related to multiple KCs. Our methods can be improved to deal with the one-to-many mapping. Besides, a more powerful BKT can be considered to incorporate the rich behavior data in MOOCs (e.g., logs of watching videos).

Acknowledgements

This paper is partially supported by the National Natural Science Foundation of China (NSFC Grant Numbers 61272343 and 61472006) as well as the National Basic Research Program (973 Program No. 2014CB340405).

References

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

[2] Matsuda, N., Furukawa, T., Bier, N., and Faloutsos, C. Machine beats experts: Automatic discovery of skill models for data-driven online course refinement. In *Educational Data Mining* (2015), 101–108.

[3] Pardos, Z., Bergner, Y., Seaton, D., and Pritchard, D. Adapting Bayesian knowledge tracing to a Massive Open Online Course in edX. In *Educational Data Mining* (2013), 137–144.