Challenges on applying BKT to model student knowledge in multi-context online learning environment

Wolney de Mello Neto, Eduardo Barbosa, Nicolau Wernerck

Felipe Garcia, Leonardo Carvalho, Pedro Carvalho

Geekie, Corp.
Sao Paulo, SP, Brazil
{wolney, eduardo.barbosa, nicolau, felipe.garcia, leonardo, pedro.carvalho}@geekie.com.br

ABSTRACT

Modelling student knowledge is a big challenge for online learning environments *(OLEs)*. One of the state-of-the-art models is the Bayesian Knowledge Tracing *(BKT)*, which estimates the probability of a student having learned a knowledge concept *(KC)* based on observable item answers over time. Nevertheless, BKT is based on a few assumptions that some real-world applications often struggle not to break, such as having homogeneous items presented to students in homogeneous contexts. Amongst other challenges pointed hereby, this poster focuses on the problem of having heterogeneous learning contexts. An experiment estimates multiple sets of parameters, one per learning context. The dataset is sampled from GeekieLab, an adaptive learning platform that is being used by more than 1 million Brazilian students.

Keywords

Bayesian Knowledge Tracing, Student Knowledge Modelling, Online Learning Environments, Adaptive Learning

1. INTRODUCTION

An online learning environment can be defined as a place where students can interact with content and/or people in order to achieve learning goals such as diagnosing knowledge gaps, learning new KCs and practicing those already known.

One major challenge in an OLE is how to measure the latent proficiency of each student in a given KC at some point in time. One could try to measure whether the student learned a KC or not, while other could try to measure how much the student knows of it. In both cases, the proficiency model should be continuously updated, that is, every interaction between the student and the platform might reflect on his proficiency, and most recent observations should have a stronger effect on calculating it. Finally, this model should estimate the probability of a student answering correctly the next item from some KC.

The BKT model[2] complies with the described requirements. Nevertheless, it comes with the expense of holding a few strong assumptions such as: (i) having a fine grained curriculum with KCs as specific as possible and dense answers data for each of them; (ii) providing homogeneous items that are related to only one specific KC; and (iii) collecting student answers from within **homogeneous learning contexts**, among many other. These are a few challenges that real-world tutoring systems face. This poster presents a brief discussion on (iii) and how online environments might not be able to hold this assumption.

2. HETEROGENEOUS CONTEXTS

The contextual effect on the BKT model has been broadly discussed in other studies, such as in [1] and [3]. In this poster, contexts refers to heterogeneous environments focusing on different stages of student learning, such as diagnosing, teaching and reinforcing.

The study case chosen for this paper is the online learning environment of GeekieLab[4], an online adaptive learning platform used by 1+ million Brazilian students. GeekieLab is a good example to illustrate the challenge discussed in this poster, since it is comprised of heterogeneous contexts where a student can answer to items. Some of its contexts are the following:

C_1 **lecture** short questions alternated with videos and slides;
C_2 **exercise list** set of questions without deadline, mainly for practice purposes;
C_3 **assessment** set of questions with short time-to-live. Exam-like environment accounted for grading.

2.1 Why should parameters be different?

As described in [2], BKT model estimates the probability $p(L)$ of a student having learned a KC by observing student answers based on a set P of the following parameters(probabilities):

$p(L0)$**:** student having learned a KC a priori;
$p(Transition)$**:** transition from unlearned to learned a KC between observations;
$p(Guess)$**:** unlearned state, but answer is right;
$p(Slip)$**:** learned state, but answer is wrong.

At first, a BKT application would estimate only one set $P_{C_{1,2,3}}$ of these 4 parameters for all observations of some KC. However, it looks more reasonable to update $p(L)$ with specific sets of parameters values (P_{C_1}, P_{C_2} and P_{C_3}) while observing item answers collected from their respective contexts C_1, C_2 and C_3. In order to investigate that, we will estimate parameters for each of the 3 contexts by training the model only with their respective observations.

3. EXPERIMENT

This experiment focuses on analyzing how parameters might vary among contexts. Upfront, the following hypotheses are proposed for further reflection:

- H_1 regarding $p(S)$, it may get higher in C_3 since items get more tricky and due to the pressure of an assessment;
- H_2 concerning $p(G)$, it may get higher in the case of a context where items provide easily detectable distractors. It could also increase in case one could have just watched a video or hint in C_1;
- H_3 $p(T)$ might be higher for C_1 since a student is presented with an item resolution/hint in between two items. Moreover, assuming a student is supposed to learn a KC before the assessment (C_3), $p(T)$ might be lower within the latter;
- H_4 assuming students face the contexts in the sequence C_1 lecture, C_2 exercise list and C_3 assessment, we expect to have an increasing $p(L0)$ throughout these them.

Based on those 3 contexts, a simple experiment was run aiming at testing the previously stated hypotheses. Further information on the data sample, estimation algorithm and method can be found in the following subsections.

3.1 Material

A sample has been extracted from GeekieLab[4]. 4 KCs were selected, each from a different domain field (Math, Portuguese Language, Natural Sciences and Human Sciences). For each of these 4 we have a data sample of 100k answers, in average $\sim$3 answers per student, and 1 answer per item. Parameters estimation were run with BKT Brute Force algorithm shared by authors from [1].

3.2 Method

For each of the 4 KCs, its $\sim$100k observations were divided into C_1, C_2 and C_3. BKT parameters were estimated per context and a full training set containing data from all contexts $C_{1,2,3}$ was used for training another set of parameters. All the search space, for earch parameter, is discretized by 0.01. Only $p(S)$ and $p(G)$ are upper-bounded by 0.1 and 0.3, respectively.

4. RESULTS AND DISCUSSION

Table 1 presents the results for estimating parameters per KC per context.

Results seem to be inconclusive for evaluating H_1 and H_2. On most scenarios, $p(S)$ and $p(G)$ are getting to the upper bound defined by the brute force implementation, intended to avoid model degeneracy. There are many possible causes to this symptom, such as noisy data from students answering without thinking fastidiously or items with easily recognizable distractors.

H_3 is somehow reflected in lecture (C_1) for KC_1 and KC_3, although KC_2 does not indicate the same. KC_4 can be discarded since there was only one item answer per student within C_1. This analysis draws attention to how important it is to filter answers for a KC from a student without some minimum number of answers, in case there should be some minimum value.

	context	p(L0)	p(T)	p(S)	p(G)	obs.
KC_1	C_1	0.06	0.20	0.1	0.3	36k
	C_2	0.46	0.03	0.1	0.3	36k
	C_3	0.34	0.15	0.1	0.20	36k
	$C_{1,2,3}$	0.27	0.13	0.1	0.29	109k
KC_2	C_1	0.77	0.041	0.1	0.3	36k
	C_2	0.33	0.001	0.1	0.3	35k
	C_3	0.33	0.231	0.1	0.3	37k
	$C_{1,2,3}$	0.58	0.001	0.1	0.3	109k
KC_3	C_1	0.73	0.09	0.1	0.3	36k
	C_2	0.46	0.02	0.1	0.3	35k
	C_3	0.13	0.08	0.1	0.3	36k
	$C_{1,2,3}$	0.47	0.04	0.1	0.3	107k
KC_4	C_1	0.79	0.001	0.02	0.16	8k
	C_2	0.51	0.18	0.1	0.3	36k
	C_3	0.21	0.54	0.1	0.3	21k
	$C_{1,2,3}$	0.54	0.24	0.1	0.3	65k

Table 1: BKT parameters estimation per context(C) for each knowledge concept(KC).

H_4 was rejected. Actually, it seems to be the contrary for KC_2, KC_3 and KC_4. C_1 has higher $p(L0)$ than C_2, which in turn has higher $p(L0)$ than C_3. This might be caused since students have better performance in C_1, tending to allow $p(L0)$ increase, indicating that the student already learned some KC.

A future investigation concerning this experiment scope could involve defining some heuristic for filtering and preprocessing the training dataset and executing this analysis on a larger dataset, with more answers per student, in order to achieve better results.

5. CONCLUSION

In general, results show that every context might need an exclusive set of parameters. In order reinforce this conclusion, an extension to the current experiment would be to evaluate model accuracy on estimating the correctness of the next answer for estimated parameters per context. This was left out of this poster due to size constraints.

All conclusions made hereby are based on simple criteria, but they manage to illustrate one of the challenging questions that one might face when implementing BKT.

6. REFERENCES

[1] R. d. Baker, A. Corbett, and V. Aleven. More accurate student modeling through contextual estimation of slip and guess probabilities in bayesian knowledge tracing. *Intelligent Tutoring Systems*, 2008.

[2] A. T. Corbett and J. R. Anderson. Knowledge tracing: Modeling the acquisition of procedural knowledge. *User modeling and user-adapted interaction*, 4(4):253–278, 1994.

[3] R. S. d Baker, A. T. Corbett, S. M. Gowda, A. Z. Wagner, B. A. MacLaren, L. R. Kauffman, A. P. Mitchell, and S. Giguere. Contextual slip and prediction of student performance after use of an intelligent tutor. pages 52–63, 2010.

[4] Geekie. Geekie lab. available at http://www.geekielab.com.br/, May 2014.