Degeneracy in Student Modeling with Dynamic Bayesian Networks in Intelligent Edu-Games

Alireza Davoodi
Department of Computer Science
The University of British Columbia
2366 Main Mall, Vancouver, BC, V6T1Z4, Canada
+1 (604) 8225108
davoodi@cs.ubc.ca

Cristina Conati
Department of Computer Science
The University of British Columbia
2366 Main Mall, Vancouver, BC, V6T1Z4, Canada
+1 (604) 8224632
conati@cs.ubc.ca

ABSTRACT
This paper investigates the issue of degeneracy in student modeling with Dynamic Bayesian Network in Prime Climb, an intelligent educational game for practicing number factorization. We discuss that maximizing the common measure of predictive accuracy (i.e. end accuracy) of the student model may not necessarily ensure trusted assessment of learning in the student and that, it could result in implausible inferences about the student. An approach which bounds the parameters of the model has been applied to avoid the issue of degeneracy in the student model to a high extent without significantly diminishing the predictive accuracy of the student model.

Keywords
Educational Games, Student Model, Dynamic Bayesian Networks, Predictive Accuracy, Model Degeneracy

1. INTRODUCTION
Assisting individuals to acquire desired knowledge and skills while engaging in a game, distinguishes digital educational games (henceforth edu-games) from traditional video games [1, 2]. Edu-games integrate game design methods with pedagogical techniques in order to more appropriately address the learning needs of the new generation, which highly regards "doing rather than knowing". Adaptive edu-games as a sub-category of edu-games leverage a user model to track the evolution of knowledge in the students and support tailored interactions with the player and have been proposed as an alternative solution for the one-size-fits-all approach used in designing non-adaptive edu-games [2].

Prime Climb (PC) is an adaptive edu-game for students in grades 5 and 6 to practice number factorization concepts. It provides a test-bed for conducting research on adaptation in edu-games. Prime Climb uses Dynamic Bayesian Network (DBN) to construct a student model which maintains and provides an assessment of student's knowledge on target skills (number factorization skills) during and at the end of the interaction. The model's assessment of the student's knowledge on the desired skills during the game play is leveraged by an intelligent pedagogical agent which applies a heuristic strategy to provide the student with personalized supports in the form of varying types of hints [3]. In addition, the model's evaluation of the student's knowledge on target skills at the end of the game, provides predictions of the student's performance on related problems outside the game environment (for instance on a post test). Therefore, an accurate student model is the main component of a system which adapts to users and any issue which could decay the efficiency of the model should be appropriately avoided and resolved.

While most of the work on user modeling in educational systems has been on optimizing the predictive accuracy (predicting student's performance on opportunities to practice skills) of the student models [5], there is limited work on educational implications and conceptual meaning imposed by the student model resulted from the predictive accuracy optimization process. This paper investigates the issue of degeneracy in the student model in PC and how it impacts the modeling. The issue of degeneracy is defined as a situation in which the parameters of a parametric student model are estimated such that the model has the highest performance (is at its global maximum given the performance and limitations of the optimization method) with respect to some standard measures of accuracy, yet it violates the conceptual assumptions (explained later in more details) underlying the process being modeled [6].

2. RELATED WORK
Difficulties in inferring student knowledge have been recently studied [4, 6, 8, 9, 10, 11] in an approach to educational user modeling called Knowledge Tracing (KT) [7]. Knowledge Tracing assumes a two-state learning model in which a skill is either in the learned or unlearned state. An unlearned skill might change to the state of learned at each opportunity the student practices the skill. In KT, it is also assumed that the student's correct/incorrect performance in applying a skill is the direct consequence of the skill being in the learned/unlearned state; yet there is always the possibility of a student correctly applying a rule without knowing the corresponding skill. This is referred to as probability of guessing. Similarly, the likelihood of a student showing an incorrect performance on applying a rule while knowing the underlying skill is called the probability of slipping. One issue with KT, called Identifiability was addressed by Beck [4]. The issue of Identifiability refers to the existence of multiple equally good mappings from observable student's performance to her corresponding latent level of knowledge while each mapping claims differently about the student performance and knowledge. To address this issues, Beck introduced the Dirichlet prior approach [4] in which a Dirichlet probability distribution is defined over the model's parameters in a KT to bias the estimation of the model parameters toward the mean of the distribution. The Dirichlet prior approach was then extended and the Multiple Dirichlet Prior approach [8] and Weighted Dirichlet Prior [9] were proposed to further address the Identifiability issue in KT. Backer et al. [6] discussed that the Knowledge Tracing models may also suffer from the problem of degeneracy. A KT model is degenerate if it updates the probability of a student knowing some skills in such a way that it violates the conceptual assumptions (such as a student being more likely to make a

correct answer if she does not have the corresponding knowledge than she does) underlying the process being modeled. Generally when the probability of slipping and guessing in KT are greater than 0.5 the model is said to be theoretically degenerate. It was also shown that the Dirichlet prior KT model (which was proposed to address the Identifiability problem in KT) also suffers from the degeneracy problem [4]. One straightforward approach to avoiding theoretical degeneration is bounding the Knowledge Tracing model parameters (probability of guessing and slipping) to take a value less than 0.5. This approach is called Bounded KT [4]. A KT model could be also empirically degenerate even if not theoretically degenerate. Two tests were also introduced to investigate empirical degeneracy in KT [4]. Baker et al. [7, 11, 12] also proposed an approach called Contextual Guess and Slip in Knowledge Tracing for contextually estimating the probabilities of guessing and slipping and showed that such model is less degenerated than standard KT which allows any value between 0 and 1 for guessing and slipping.

This paper builds on the previous works on issues with Knowledge Tracing, to investigate the issue of degeneracy in a student model which uses a Dynamic Bayesian Network and a causal structure to infer about the student's knowledge on skills in an adaptive edu-game called Prime Climb. The issues of degeneracy has been studied in Knowledge Tracing models which assume that learning different skills is independent from each other while in PC, based on guidance from a math expert, it is assumed that the factorization skills are not independent from each other. Moreover, In KT, at each time, the student has an opportunity to practice a single skill, while in Prime Climb, at least three skills are practiced simultaneously and consequently there are other model's parameters than probability of guessing and slipping in the student model in PC.

3. PRELIMINARIES/BACKGROUND

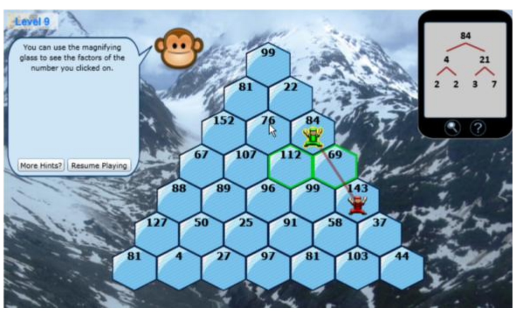


Figure 1: Prime Climb Edu-game

3.1 Prime Climb Edu-game:

In Prime Climb (Figure 1), the player and her partner climb a series of mountains (11 mountains) of numbers by pairing up the numbers which do not share a common factor. The main interaction of a player with Prime Climb consists of making a movement from a location on a mountain of numbers to another location on the mountain until she reaches the top of the mountain. Therefore at each movement, the student practices at least 3 skills: 1) Factorization of the number the player moves to. 2) Factorization of the number the partner is on and 3) The concept of common factor between the 2 numbers.

3.2 Student Model in Prime Climb:

Prime Climb is equipped with 11 probabilistic student models (one for each mountain) which use Dynamic Bayesian Network to model the evolution of student's factorization knowledge during the period of time that she interacts with Prime Climb. To this end the student model consists of time slices representing relevant temporal states in the process being modeled. Each time slice is created once a student makes a movement (climbs a mountain). The smallest student model in PC consists of 23 binary nodes (random variables) and the largest one contains 131 nodes.

<u>PC's Student Model Nodes</u>: In PC, each student model contains several binary nodes [5] such as:

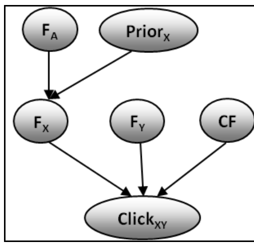
Factorization Nodes (F_X): Each factorization node, F_X, is a binary random variable which represents the probability that the student has mastered the factorization skill of number X.

Common Factor Node (CF): There is only one CF node. It is a binary random variable representing the probability that student has mastered the concept of common factor between numbers.

$Prior_X$ Node: There is one Prior node for each none-root factorization node in the model. It shows the prior probability that the student knows the factorization of the number X to its factors.

Click Nodes ($Click_{XY}$): Once the player makes a move (i.e. moves to number X while the partner is on Y) a Click node is temporarily added as a child of the three random nodes F_X, F_Y and CF to make a causal structure. Therefore, these three nodes are conditionally dependent to each other given evidence on the Click node. Such causal structure allows apportion of blame for wrong movements [5]. Table 1 and Table 2 show the Conditional Probability Table (CPT) of the F_X and Click nodes respectively.

Table 1: Model Structure and CPT of F_X Factorization Node



F_A	$Prior_X$	$P(F_X = Known)$
Known	Known	1
Known	Unknown	$Max \times \frac{P_k}{P}$
Unknown	Known	1
Unknown	Unknown	0

Table 2: CPT of Click (K: Known, U: Unknown, C: Correct)

	$F_X = K$		$F_X = U$		
	F_Y=K	F_Y=U	F_Y=K	F_Y=U	
$P(Click_{XY}=C)$	1-Slip	Edu-Guess	Edu-Guess	Guess	CF=K
	Guess	Guess	Guess	Guess	CF=U

<u>Model's Parameters in Prime Climb</u>: The parameters (guess, edu-guess, slip and max) are called "model's parameters" in the student model in Prime Climb:

Slip: The probability of making a wrong action on a problem step when the student has the corresponding knowledge.

Guess: The probability of making a correct action on a problem step when the student does not have the corresponding skill.

Edu-Guess: The probability of a student making a correct answer while the student does not completely master the required knowledge for making such correct action.

Max: A coefficient in the formula ($Max \times \frac{P_k}{P}$) used to calculate the probability of a student making a correct move proportional to number of its known parents. (P is number of parents of F_X and P_K is number of those parents which are known.

<u>End Accuracy of Student Model</u>: The student model in PC is evaluated based on the end accuracy (=predictive accuracy) of the model. The end accuracy is defined as the model's performance in accurate assessment of the student's factorization knowledge about some sample numbers appearing on a post-test at the end of the game and calculated using the following formula [13]:

$$End\ Accuracy = \frac{Sensitivity + Specificity}{2}$$

Before starting the game, the factorization and common factor nodes in the student model are initialized with prior probabilities that the student knows the number factorization and common factor concept. In Prime Climb, three types of prior probability are used which are defined as following:

Generic: The prior probability that a student knows number factorization and common factor skills is set to 0.5.

Population: The prior probability is calculated based on scores of a group of students on a pre-test which examines the knowledge of students on specific factorization skills.

User-specific: The prior probability is specific to each student based on her performance on the pre-test. If a student has correctly responded a number factorization question in the pre-test, the probability that the student knows the corresponding factorization skill is set to 0.9 otherwise it is set to 0.1.

Plausibility of Parameters in PC's Student Model: While the end accuracy is used to evaluate PC's student model, the model's parameters can also be directly evaluated based of "plausibility" criteria. One criterion is the impact of model's parameters (guess, edu-guess, slip and max) on performance of the adaptive interventions (hints) mechanism in PC. The performance of hinting mechanism in PC is calculated based on average of two F-Measures [12]: 1) Positive F-Measure: calculated using precision and recall of the hinting mechanism in identifying correct time points for providing hints and 2) Negative F-Measure: calculated using precision and recall of the hinting mechanism in identifying the time points in which hints should not be given to the users. According to such criteria, a set of model's parameters improving performance of the hinting mechanism while providing reasonable number of hints during game play is more plausible. For instance if a student makes 200 movements in total during the game, it is not plausible to receive over 100 hints (One hint for every two movement on average). Notice that the value of the model's parameters directly affects the hinting mechanism in PC.

4. DEGENERACY IN STUDENT MODEL

Student Model Optimization: The Prime Climb's original student model allows any value between 0 and 1 for the model's parameters (slip, guess, edu-guess and max). The values for the parameters are estimated such that the model's end accuracy is maximized. To this end, an exhaustive search procedure is applied which examines values between 0 and 1 in interval of 0.1 for each parameters and eventually selects the parameters combination maximizing the end accuracy. To this end, a Leave-One-Out Cross Validation approach was applied across 43 students who played Prime Climb. The optimal set of parameters and the mean end accuracy across the test folds for each prior probability type were computed and summarized in Table 3.

Table 3: Estimated Parameters and End Accuracy

Prior Probability	Guess	Edu-Guess	Max	Slip	End Accuracy (M/SD)
Population	0.5	0.3	0.2	0.4	0.77/0.14
Generic	0.2	0.6	0.8	0.6	0.70/0.15
User-specific	0.6	0.1	0.6	0.6	0.72/0.20

Degeneracy in Student Model: The degeneracy in student modeling in Prime Climb is defined as violation of the conceptual assumptions behind modeling of a student's knowledge on factorization skills during interaction with the game. The conceptual assumptions in PC student model are as following:

1) Correct evidence (action) on a skill must not decrease the probability of the student knowing the skill.
2) An incorrect action on a skill must not increase the probability of the student knowing the skill.

Any pattern in the student model violating the aforementioned assumptions is marked as model degeneration. We defined two tests to investigate model degeneracy in Prime Climb's student model. If the student model fails either of these two tests, the model is said to be degenerated:

Test 1 of degeneration in Prime Climb: If a student makes a correct/incorrect action on an opportunity to practice a skill, the probability of the student knowing the skill should not be less/greater than the probability of knowing the skill before making the action on the skill. Mathematically the following cases show failures in Test 1:

$$P(F_X = Known\ |Click = Correct) < P(F_X = Known)$$
$$P(F_Y = Known\ |Click = Correct) < P(F_Y = Known)$$
$$P(F_X = Known\ |Click = Wrong) > P(F_X = Known)$$
$$P(F_Y = Known\ |Click = Wrong) > P(F_Y = Known)$$
$$P(CF = Known\ |Click = Correct) < P(CF = Known)$$
$$P(CF = Known\ |Click = Wrong) > P(CF = Known)$$

Test 2 of degeneration in Prime Climb: Assume a dependency relationship between two skills S_1 and S_2 such that knowledge on S_1 implies knowledge on S_2 with a certain probability. If a student performs correctly/incorrectly on an opportunity to practice skill S_1, the probability that the student knows skill S_2 should not be less/greater than its values before making the action.

The original student model was checked for degeneracy using the Tests 1 and 2 of degeneration. Table 4 summarizes the mean number of failures across the 43 students who played PC.

Table 4: Failures in Test 1 and Test 2 in PC's Original Model

Prior Probabilities	Failures in Test 1 (M/SD)	Failures in Test 2 (M/SD)
Population	268.91/64.26	1.71/2.85
Generic	101.84/28.73	258.35/80.48
User-specific	339.17/75.11	138.86/62.04

As shown in Table 4, the original student model in Prime Climb suffers from degeneracy issue. Theoretically, based on the CPT of the Click node (See Table 2), it can be concluded that the following conditions (in Table 5) might cause specific patterns of degeneracy in the Prime Climb's student model.

Table 5: Conditions and Patterns of Degeneracy in PC

Conditions	Related Patterns of Degeneracy
Eduguess< Guess	$P(CF = Known\ \|Click = Correct) < P(CF = Known)$ $P(CF = Known\ \|Click = Wrong) > P(CF = Known)$
1-Slip < Guess	
1-Slip < Eduguess	$P(F_Y = Known\ \|Click = Correct) < P(F_Y = Known)$ $P(F_Y = Known\ \|Click = Wrong) > P(F_Y = Known)$ $P(F_X = Knwon\ \|Click = Correct) < P(F_X = Known)$ $P(F_X = Known\ \|Click = Wrong) > P(F_X = Known)$

Given the estimated parameters for the original presented in Table 3 and the degeneracy conditions in Table 4, different patterns of degeneracy can be observed in the Prime Climb's original model.

5. BOUNDED STUDENT MODEL

To alleviate the issue of degeneracy, the model's parameters are bounded to take values from outside the subspaces (conditions in Table 5) that cause specific patterns of degeneracy in PC. Such model is called Prime Climb's Bounded student model. Similar to the PC's original model, an exhaustive search approach is used to find a set of bounded model's parameters which maximizes the model's end accuracy. In this study we allow values greater than 0.5 for the model's parameters. The estimated parameters and the end accuracy of the bounded student model are shown in Table 6.

Table 6: The Estimated Parameters and End Accuracy

Prior Probability	Guess	Edu-Guess	Max	Slip	End Accuracy (M/SD)
Population	0.7	0.7	0.2	0.4	0.76/0.15
Generic	0.5	0.6	0.4	0.8	0.68/0.15
User-specific	0.3	0.3	0.6	0.8	0.70/0.18

Comparison of the Models' Accuracy: The results of a paired t-test showed no statistically significant difference between the end accuracy and AUC (Area under the ROC curve) of the original and bounded models in none of the prior probability type.

Table 7: AUC of the student models

AUC	Prior probability Types		
	Population	Generic	User-specific
Original	0.7345	0.6762	0.7860
Bounded	0.7375	0.6643	0.7449

Comparison of Models' Degeneracy: A paired t-test is used to compare the two models based on the average number of failures in the two tests of degeneracy. Table 8 shows the results. In all cases, the bounded model resulted in significantly lower number of failures in the both tests of degeneration and the p-value is less than 0.01 (except where indicated by *).

Table 8: Comparison of Failures in Degeneration Tests

Tests	Models	Population (Mean/SD)	Generic (Mean/SD)	User-specific (Mean/SD)
Test 1	Bounded	9.17 / 7.45	8.8/7.29	10.24/10.96
	Original	268.91 / 64.26	101.84/28.73	339.17/75.11
Test 2	Bounded	1.44 / 2.0	0.24/0.48	0.53/1.2
	Original	1.71 / 2.86*	258.35/80.48	138.86/62.04

Comparison of the Models' Parameters Plausibility: The plausibility of the estimated parameters in original and bounded models was compared based on performance (measured by F-Measure as described before) of the hinting method in PC. To this end, the performance of the hinting mechanism as well as average number of given hints are calculated. A paired t-test is used to compare the hinting procedure performance and number of hints across 43 students. The following tables show the comparison results. In all comparisons, the p-value is less than 0.01.

Table 9: Comparison of F-Measures of Hinting Mechanism

F-Measure	Population (Mean/SD)	Generic (Mean/SD)	User-specific (Mean/SD)
Original	0.24 / 0.2	0.3 / 0.24	1 / 0
Bounded	0.29 / 0.22	0.55 / 0.32	0.95 / 0.08

Table 10: Comparison of number of adaptive hints

#Hints	Population (Mean/SD)	Generic (Mean/SD)	User-specific (Mean/SD)
Original	112.5 / 56.62	82.95 / 27.37	139 / 39.36
Bounded	55.2 / 19.84	48.53 / 19.1	42 / 18.24

As shown in Table 10, the results of a paired t-test show that the bounded models resulted in a significantly lower number of hints ($p<0.01$ in all cases) while significantly higher performance for the hinting mechanism (except for the student model with user-specific prior probability type). Note that on average each student makes 164.5 movements while playing PC. Based on the results, the hinting mechanism provides 2, 1.7 and 3.3 times more hints in the original model than the bounded with population, generic and user-specific prior probability types respectively. This shows that in general, the bounded model provides more plausible model's parameters than the original student model.

6. CONCLUSIONS/FUTURE WORK

This paper discussed that optimizing the student model in Prime Climb does not ensure a trusted student modeling because the model might be degenerated. The issue of degeneracy and sources and patterns of degeneracy were described and one approach to addressing this issue called, bounded model was also introduced and compared with the original student model. It was shown that the bounded model has a comparable accuracy with the original model while it contains significantly fewer cases of degeneracy. The estimated parameters in the bounded model were also more plausible than the parameters in the original model. In the current bounded model, the model's parameters are estimated the same across all students. As for future work, we will consider more personalized model's parameters in bounded model to account for individual differences between users.

7. REFERENCES

[1] de Castell, S. & Jenson, J., 2007, Digital Games for Education: When Meanings Play. Intermedialities, 9, 45-54.

[2] Conati, C. and M. Klawe, 2002, Socially Intelligent Agents in Educational Games. In Socially Intelligent Agents - Creating Relationships with Computers and Robots. K. Dautenhahn, et al., Editors, Kluwer Academic Publishers.

[3] Conati C and Manske M.: Evaluating Adaptive Feedback in an Educational Computer Game, IVA 2009, 146-158

[4] Beck, J.E., 2007, Difficulties in inferring student knowledge from observations (and why you should care). Educational Data Mining

[5] Manske, M., Conati, C., Modelling Learning in an Educational Game. AIED 2005: 411-418

[6] Baker, R. S.J.d., Corbett, A.T., Aleven, V., 2008, More Accurate Student Modeling Through Contextual Estimation of Slip and Guess Probabilities in Bayesian Knowledge Tracing. Human-Computer Interaction Institute. Paper 6. http://repository.cmu.edu/hcii/6

[7] Corbett, A.T. and Anderson, J. R., 1995, Knowledge tracing: Modeling the acquisition of procedural knowledge, User Modeling and User Adapted Interaction, Volume 4, Number 4, 253-278

[8] Gong, Y., Beck. J. E., Ruiz, C., 2012, Modeling Multiple Distributions of Student Performances to Improve Predictive Accuracy. UMAP 2012: 102-113

[9] Rai, D., Gong, Y., Beck, J., 2009, Using Dirichlet priors to improve model parameter plausibility. EDM 2009: 141-150

[10] Baker, R. S.J.d., Corbett, A.T., Aleven, V., 2008, Improving Contextual Models of Guessing and Slipping with a Trucated Training Set. EDM 2008: 67-76

[11] Baker, R. S.J.d., Corbett, A.T., Gowda, S. M., Wagner, A.Z., MacLaren, B. A., Kauffman, L. R., Mitchell, A. P., Giguere, S., 2010, Contextual Slip and Prediction of Student Performance after Use of an Intelligent Tutor. UMAP 2010: 52-63

[12] Powers, David M W, 2011. "Evaluation: From Precision, Recall and F-Factor to ROC, Informedness, Markedness & Correlation". Journal of Machine Learning Technologies **2** (1): 37–63.

[13] Altman DG, Bland JM (1994). "Diagnostic tests. 1: Sensitivity and specificity". *BMJ* 308 (6943): 1552