Defining Mastery: Knowledge Tracing Versus N- Consecutive Correct Responses

Kim Kelly
Worcester Polytechnic Institute
100 Institute Road
Worcester, MA 01609
508-461-6386
kkelly@wpi.edu

Yan Wang &
Tamisha Thompson
Worcester Polytechnic Institute
100 Institute Road
Worcester, MA 01609
Ywang14@wpi.edu

Neil Heffernan
Worcester Polytechnic Institute
100 Institute Road
Worcester, MA 01609
508-831-5569
nth@wpi.edu

ABSTRACT

Knowledge tracing (KT) is well known for its ability to predict student knowledge. However, some intelligent tutoring systems use a threshold of consecutive correct responses (N-CCR) to determine student mastery, and therefore individualize the amount of practice provided to students. The present work uses a data set provided by ASSISTments, an intelligent tutoring system, to determine the accuracy of these methods in detecting mastery. Study I explores mastery as measured by next problem correctness. While KT appears to provide a more stringent threshold for detecting mastery, N-CCR is more accurate. An incremental efficiency analysis reveals that a threshold of 3 consecutive correct responses provides adequate practice, especially for students who reach the threshold without making an error. Study II uses a randomized- controlled trial to explore the efficacy of various N-CCR thresholds to detect mastery, as defined by performance on a transfer question. Results indicate that higher thresholds of N-CCR lead to more accurate predictions of performance on a transfer question than lower thresholds of N-CCR or KT.

Keywords

Intelligent Tutoring System, Knowledge Tracing, Mastery Learning.

1. INTRODUCTION

Intelligent tutoring systems are known for their ability to personalize the learning experience for students. One way that learning is individualized is by providing just the right amount of practice to meet the student's needs. Determining the correct amount of practice is critical because over-practice might bore students and take an un-necessarily long time, while under-practice might not provide enough opportunities for a student to learn a skill. To determine the correct amount of practice, systems must identify the point in time when students have learned the skill, otherwise referred to as reaching mastery.

Defining mastery may vary between systems. One measure of mastery includes next problem correctness, another is performance on a transfer question, and yet another is performance on a delayed retention test. Some systems rely on knowledge tracing (KT) [1-2], others use a predetermined number of consecutive correct responses (N-CCR) [3, 4, 9}. In each case, mastery status is used by the system to determine the end of an assignment.

2. METHODOLOGY

This research is comprised of two studies, the first was a data analysis of large data sets provided by ASSISTments, and the second was a randomized controlled trial. Study I of the present study leverages data generated by an intelligent tutoring system to explore the ability of N-CCR and KT to detect mastery. Mastery will be measured by next problem correctness. Additionally, an incremental efficiency analysis will also be presented that sheds light on the number of additional questions students must answer to reach a given threshold.

Next problem correctness is arguably a weak measure of mastery as slips are possible. A measure of more robust learning is performance on a transfer task [10]. Therefore, in Study II, a randomized-controlled trial was conducted to compare the accuracy of different potential thresholds of number of consecutive correct responses. This data was then used to further explore KT predictions, compared to N-CCR in an attempt to determine which method should be used in intelligent tutoring systems who rely on mastery to determine amount of practice.

3. RESULTS

3.1 NCCR

When mastery is defined by next problem correctness, results indicate that 3-CCR is an adequate threshold for accurately detecting mastery. Table 1 shows that 80% of students who answer three questions correctly, go on to answer the fourth and fifth correctly as well.

Table 1: Percentage of students with each response combination of the fourth and fifth question following 3-CCR.

3 Consecutive No Errors		Fourth Question	
		Incorrect	Correct
Fifth Question	Incorrect	1.8% (5)	9.8% (24)
	Correct	8.4% (28)	80.0% (228)

When mastery is defined by performance on a transfer question, results indicate that 5-CCR (Table 3) more accurately detects mastery than 3-CCR (Table 2). Accuracy is defined by the percentage of students who met the threshold and were successful on the transfer questions combined with the percentage of students who failed to meet the threshold and answered the transfer questions incorrectly. Identifying students who met the threshold yet answered the transfer incorrectly are considered false positives and students who answered the transfer question correctly yet failed to meet the threshold are considered false negatives.

Table 2: Student performance on transfer question based on 3-CCR.

Percent(Number) of students	Threshold Met	Threshold Not Met
Transfer Correct	46%(17)	0%
Transfer Incorrect	43%(16)	11%(4)

Table 3: Student performance on transfer question based on 5-CCR.

Percent(Number) of students	Threshold Met	Threshold Not Met
Transfer Correct	43%(16)	8%(3)
Transfer Incorrect	19%(7)	30%(11)

3.2 KT

When mastery is defined by next problem correctness, results indicate that KT is comparable to 3-CCR in accurately detecting mastery for students who do not make an error (Table 4).

Table 4: Accuracy of KT detecting mastery for students who answered three consecutive questions correctly without an error. (n=287)

	Threshold Met (>95%)	Threshold Not Met (<95%)
Next Question Correct	80.5% (231)	9.4% (27)
Next Question Incorrect	8.4% (24)	1.7% (5)

When mastery is defined by performance on a transfer question, results indicate that KT is comparable to 3-CCR, but less accurate than 5-CCR (Table 5).

Table 5: . Student performance on the transfer question based on KT's 95% threshold.

Percent(Number) of students*	Threshold Met	Threshold Not Met
Transfer Correct	42%(31)	7%(5)
Transfer Incorrect	39%(29)	12%(9)

3.3 Incremental Efficiency Analysis

Using the data generated from the students reaching the 5-CCR threshold, we determined how many additional questions were required to reach each incremental threshold. This provides insight into the tradeoff between potential increased mastery detection and time consumption, as measured by number of questions completed. 3-CCR is a sufficient threshold, as over 90% students go on to reach the higher threshold. Of the students who reached the final 5-CCR threshold, 90% of them reached it without an error. Those who made at least one error, tended to reach the threshold with N attempts following the error. This suggests that the error was a slip.

4. DISCUSSION

Accurately predicting or detecting mastery status is critical to intelligent tutoring systems, because the amount of practice provided to students depends on this. An overly cautious prediction will lead to unnecessary practice (false negatives), while less strict criteria will not provide enough (false positives).

N-CCR, specifically 3-CCR, is a simple, yet effective way to determine mastery within an ITS. This threshold has been found to predict next problem correctness with at least 80% accuracy. However, when predicting performance on a transfer task, a higher threshold (5-CCR) is more effective. Both thresholds of N-CCR were more accurate than the more complicated method, knowledge tracing, when determining mastery.

5. ACKNOWLEDGMENTS

We thank multiple NSF grant (ACI-1440753, DRL-1252297,DRL-1109483,DRL-1316736 & DRL-1031398) , the US Dept of Ed's (IES R305A120125 & R305C100024 and GAANN), the ONR, and the Gates Foundation.

6. REFERENCES

[1] Corbett, A., Anderson, J. (1995). Knowledge Tracing: Modeling the Acquisition of Procedural Knowledge. *User Modeling and User-Adapted Interaction, 4*, 253-278.

[2] Fanscali, Stephen E., Nixon, Tristan, & Ritter, Stephen (2013). Optimal and worst-case performance of mastery learning assessment with bayesian knowledge tracing. *Proceedings of the 6th International Conference on Educational Data Mining.* D'Mello, S., Calvo, R., Olney, A. (Eds). 35-42.

[3] Faus, M. (2014). Improving Khan Academy's student knowledge model for better predictions. *MattFaus.com* [web log]. Retrieved October, 2014, from http://mattfaus.com/2014/05/improving-khan-academys-student-knowledge-model-for-better-predictions/

[4] Feng, M., Heffernan, N. T., Koedinger, K. R.: Addressing the assessment challenge in an Intelligent Tutoring System that tutors as it assesses. *User Modeling and User-Adapted Interaction*, 19, 243-266 (2009)

[9] Hu, D. (2011). How Khan Academy is using Machine Learning to Assess Student Mastery. *David-Hu.com* [web log]. Retrieved October, 2014, from http://david-hu.com/2011/11/02/how-khan-academy-is-using-machine-learning-to-assess-student-mastery.html