The Complex Dynamics of Aggregate Learning Curves

Tristan Nixon, Stephen Fancsali, and Steven Ritter

Carnegie Learning, Inc.
437 Grant Street, Suite 918
Pittsburgh, PA 15219, USA
(888) 851.7094 {x123, x219, x122}
{tnixon, sfancsali, sritter}@carnegielearning.com

ABSTRACT

Mastery learning in intelligent tutoring systems produces a differential attrition of students over time, based on their levels of knowledge and ability. This results in a systematic bias when student data are aggregated to produce learning curves. We outline a formal framework, based on Bayesian Knowledge Tracing, to evaluate the impact of differential student attrition in mastery learning systems, and use simulations to investigate the impact of this effect in both homogeneous and mixed populations of learners.

Keywords

Mastery learning, attrition bias, learning curves, aggregate learning, heterogeneous learner populations, knowledge tracing

1. MASTERY ATTRITION BIAS

Attrition bias occurs when some aspect of an experimental design has a significant and systematic effect on whether subjects complete all measures [6]. Although students working with an intelligent tutoring system (ITS) are not *ipso facto* in any experimental conditions, the mastery learning assessment built into many such systems creates an attrition bias. ITSs that implement mastery learning assess a student's performance as she works through instructional material, and continually re-evaluate whether she has received sufficient practice on targeted skills or knowledge components (KCs). This is a commonly used method to allocate student time, but by selectively removing students who master material quickly from the sample, it differentially biases the resulting data in ways that may conceal the learning of individuals [4]. ITSs that re-visit previously mastered KCs may exhibit this same effect only within blocks of contiguous practice.

In the Bayesian Knowledge Tracing (BKT)[2] model of student learning, the performance of an individual student can be described by the equation:

$$P_c(t) = P_k(t)(1-\theta_s) + (1-P_k(t))\theta_g$$

where $P_c(t)$ is the probability that the student will give a correct response at time *t*, given the probability of student knowledge, $P_k(t)$, and the performance parameters θ_s, θ_g (slip and guess). Consider a homogenous population of learners, all with the same parameters. We describe the average correctness of responses as:

$$\bar{C}(t) = \frac{K(t) - S(t) + G(t)}{K(t) + U(t)}$$

where $K(t)$ and $U(t)$ are the numbers of students in the known and unknown states at time *t*, respectively. $S(t)$ and $G(t)$ are binomial random variables giving the numbers of slips and guesses:

$$S(t) \sim B(K(t), \theta_s), \qquad G(t) \sim B(U(t), \theta_s)$$

It can be shown that the expected behavior of the aggregate learning curve:

$$E[\bar{C}(t)] = E[\bar{K}(t)](1-\theta_s) + (1 - E[\bar{K}(t)])\theta_g$$

is controlled by the ratio of students in the known state:

$$\bar{K}(t) = \frac{K(t)}{K(t) + U(t)}$$

The aggregate learning curve may be described as a weighted average between the expected performance in the known and unknown states, weighted by the ratio of students in each. The known and unknown populations will change according to the following stochastic recurrence relations:

$$K(t) = K(t-1) + L(t) - M_k(t)$$

$$U(t) = U(t-1) - L(t) - M_u(t)$$

where $L(t)$ is the number of students who learn the skill at time *t*, and so transition from the unknown into the known state. It is also binomially distributed: $L(t) \sim B(U(t-1), \theta_l)$, where θ_l is the BKT learning (aka. transition) parameter. $M_k(t)$ and $M_u(t)$ give the numbers of students from the known and unknown states, respectively, that are judged to have mastered the material by the system, and so removed from the population. The initial share of students in the known and unknown states is controlled by θ_i, the initial knowledge parameter from BKT:

$$K(1) = I \sim B(N, \theta_i).$$

From this we can see that the learning curve begins from a theoretical initial value of:

$$E[\bar{C}(1)] = \theta_i(1-\theta_s) + (1-\theta_i)\theta_g$$

In the no-mastery attrition situation, where $M_k(t)$ and $M_u(t)$ are always 0, $\bar{K}(t)$ will tend towards 1. Therefore the learning curve will converge to a theoretical maximum:

$$\lim_{t\to\infty} E[\bar{C}(t)] = (1-\theta_s)$$

We see this behavior in the left-hand plot of Figure 1.

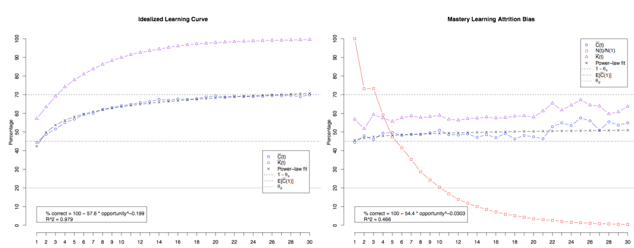

Figure 1: simulated learning curves with (right) and without (left) mastery learning

However, when mastery learning is involved, we must consider a balance of factors. We can expand the $\overline{K}(t)$ ratio in terms of its recurrence relations:

$$\overline{K}(t) = \frac{K(t-1) + L(t) - M_k(t)}{K(t-1) + U(t-1) - M_k(t) - M_u(t)}$$

Assuming $M_u(t)$ is negligible, the changes in this ratio depend on the relative magnitudes of $L(t)$ and $M_k(t)$. Except for when the unknown population has diminished to zero, the denominator of the ratio is larger than the numerator, so subtracting $M_k(t)$ from both will lead to a reduction in $\overline{K}(t)$. Since a falling $\overline{K}(t)$ ratio increases the weight that the unknown states play in the aggregate curve, mastery learning leads to lower aggregate performance, *ceteris paribus*.

Although learning and mastery have opposite effects on the instantaneous change in the $\overline{K}(t)$ ratio, they are not constant or independent over time. Learning has a negative-feedback relationship to itself: it reduces the size of the unknown student population, so the expected value of $L(t)$ will diminish over time. Mastery also has a negative-feedback relationship with the known population, but learning tends to counter-act that effect. Thus, learning has a positive-reinforcement relationship on mastery. In sum, there are many reasons why mastery learning leads to aggregate learning curves that do not take the shape we expect in their idealized form.

2. HETEROGENEOUS POPULATIONS

So far we have been considering idealized situations in which all students are instances of a BKT model with a common set of parameters. Naturally, we wish to investigate what can happen to aggregate learning curves when we have a heterogeneous population of different learners. There are very many different possible ways a heterogeneous population might be composed, and there could be very many perverse aggregate learning curves created by specially constructed mixed populations. We illustrate just a couple of examples that show interesting aggregate behavior.

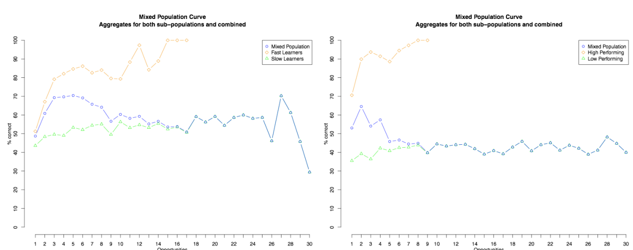

Figure 2: simulated heterogeneous populations.

Figure 2 demonstrates a couple of examples representing the range of aggregate behavior possible in mixed populations. In the left-hand plot, we show a population with similar initial knowledge, but composed of both fast-learning and slow-learning students. In the right-hand plot, we have a mixed population of higher-performing and lower-performing students. In both cases, the initial opportunities are a balanced mix of both populations. However, as the better students are preferentially removed by the mastery-learning system, they represent a diminishing fraction of the total population, and eventually the aggregate curve converges to that of the lower-performing sub-population. In the one case, the aggregate curve demonstrates a rising and falling pattern, whereas in the other case, the curve appears to demonstrate "negative learning". In a mixed population, the frequency of correct responses is the weighted average across the (j) sub-populations:

$$\bar{C}(t) = \frac{1}{N(t)} \sum_{j=1}^{J} N_{[j]}(t) \left[K_{[j]}(t) - S_{[j]}(t) + G_{[j]}(t) \right]$$

We could easily extend this notation of sub-populations to distinct per-student learning profiles. In this situation, $J = N(t)$ and $N_{[j]}(t)$ is either 1 or 0, depending on whether the j^{th} student has "mastered-out" or not.

3. CONCLUSIONS

Aggregate learning curves are used to evaluate and improve instructional systems[3]. However, there are significant distortions to aggregate measures of student learning created by the differential attrition bias inherent to mastery learning systems. Aggregate performance on each step shown by learning curves need not be representative of the learning of individuals or groups of students [4]. Aggregate measures of such attrition-biased data will tend to under-represent the amount of learning occurring. Explicitly modeling the effect of this attrition bias may be a fruitful direction for future research.

A mixed population with different learning characteristics can introduce additional distortions when mastery learning is involved. There has been much work already on identifying the learning characteristics of individuals and sub-populations[1][5]. Further developments in this direction would help build richer and more accurate models of learning robust to the attrition bias in mixed-population data.

4. REFERENCES

[1] Baker, R. Corbett, A. T., Aleven, V. 2008. More Accurate Student Modelling through Contextual Estimation of Slip and Guess Probabilities in Bayesian Knowledge Tracing. *Proceedings of the 9th International Conference on Intelligent Tutoring Systems*. ITS 2008. 406-415

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

[3] Martin, B., Mitrovic, A., Koedinger, K. R., Mathan, S., 2011. Evaluating and improving adaptive educational systems with learning curves. *User Modeling and User-Adapted Interaction*. 21, 3, 249-283

[4] Murray, C., Ritter, S., Nixon, T., Schwiebert, R., Hausmann, R., Towle, B., Fancsali, S., Vuong, A. 2013. Revealing the Learning in Learning Curves. *Proceedings of the 16th International Conference on Artificial Intelligence in Education* (Memphis, TN, July 9-13 2013). AIED 2013.

[5] Pardos, Z. A., Heffernan, N. T. 2010. Modeling individualization in a Bayesian networks implementation of knowledge tracing. *Proceedings of the 18th International Conference on User Modeling, Adaptation and Personalization.* 255-266

[6] Shadish, W., Cook, T., Campbell, D. 2002. *Experimental and Quasi-Experimental Designs for Generalized Causal Inference.* Houghton Mifflin Company, Boston, MA.