Recent-Performance Factors Analysis

April Galyardt
University of Georgia
110 Carlton St.
Athens, GA
galyardt@uga.edu

Ilya Goldin
Center for Digital Data, Analytics, and Adaptive Learning
Pearson
ilya.goldin@pearson.com

ABSTRACT
We introduce R-PFA, a new model for predicting whether or not a student will answer an item correctly based on the student's history of practice. The key idea in R-PFA is to represent history as a recentcy-weighted proportion of correct responses. In an evaluation on a dataset from the Assistments tutoring system, we find that R-PFA improves predictive accuracy over other logistic regression model variants, including PFA and AFM.

Keywords
performance modeling, moment of learning, linear logistic test model

1. INTRODUCTION
An interactive learning environment (ILE) can adapt its behavior to what the student does and does not know. For example, an ILE may hold a domain model in terms of knowledge components (KCs) to be taught to students [5], and estimate each student's proficiency with a KC based on the student's practice with problems involving the KC. One popular model for estimating student proficiency in this setting is Performance Factors Analysis (PFA) [6]. PFA is a parameterization of the Linear Logistic Test Model [3] that predicts performance on the current item using the entire history of success and failures on previous items addressing the same KC (Eq 1).

We introduce Recent-Performance Factors Analysis (R-PFA), which extends PFA via a simple variable transformation. R-PFA includes information about whether or not a moment of learning has occurred [4]. We demonstrate that R-PFA shows improved predictive performance over PFA as well as over the Additive Factors Model (AFM) [2]. We also contend that the simplicity of R-PFA provides it with a distinct advantage over the methodology in [4] for prediction.

2. METHODS
R-PFA is based on a simple observation: If a student has already experienced a moment of learning then recent performance is likely to consist primarily of successful attempts. If a moment of learning has not yet occurred, then recent performance is likely to contain more failed attempts. In effect, recent history serves as a proxy for whether or not a moment of learning has occurred.

We compare the performance of R-PFA (Eq 2) to PFA (Eq 1), AFM, and a simplified version of PFA using success only, on the Assistments data used in the original "moment of learning" work [1]. The data contain first attempts by 4138 students on problem sets involving 54 knowledge components (KC), for a total of 187,309 first attempts. Each problem is coded with only a single KC. Table 1 lists the features in each model.

$$logit(p_{ijt}) = \theta_i + \beta_j + \alpha_j S_{ijt} + \rho_j F_{ijt} \tag{1}$$

$$logit(p_{ijt}) = \theta_i + \beta_j + \gamma_j T_{ijt} + \delta_j R_{ijt} \tag{2}$$

We use the notation:

j	KC indicator
i	student indicator
X_{ijt}	binary correct/incorrect, student i, KC j, trial t
S_{ijt}	count of previous successes, up to trial t
F_{ijt}	count of previous failures, up to trial t
T_{ijt}	count of past opportunities, $S_{ijt} + F_{ijt}$
R_{ijt}	recency-weighted proportion of past successes
p_{ijt}	$Pr(X_{ijt} = 1)$.

Table 1: Terms in predictive model variants.

	Student	KC	Success	Failure	Totals	Weighted Proportion
AFM	θ_i	β_j			$\gamma_j T_{ijt}$	
sPFA	θ_i	β_j	$\alpha_j S_{ijt}$			
PFA	θ_i	β_j	$\alpha_j S_{ijt}$	$\rho_j F_{ijt}$		
R-only	θ_i	β_j				$\delta_j R_{ijt}$
R-PFA	θ_i	β_j			$\gamma_j T_{ijt}$	$\delta_j R_{ijt}$

As a measure of 'recent' history, we introduce R_{ij}, an exponentially weighted proportion of successes.

$$R_{ijt} = \frac{\sum_{p=-2}^{t-1} b^{(t-p)} X_{ijp}}{\sum_{p=-2}^{t-1} b^{(t-p)}} \tag{3}$$

The decay factor b is a tuning parameter that controls the weights, and thus controls whether 'recent' means just the most recent trial or the entire history of practice. We compare values of {0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0} for b.

One potential issue with R_{ij} is that the proportion of recent successes is very noisy on the first few trials. To adjust for this noise, we stipulate ghost attempts $X_{ij-2} = X_{ij-1} = X_{ij0} = 0$. The ghost attempts are an explicit assumption that at time 0, the student has not already learned the KC. These ghost attempts affect only the value of R_{ijt}, i.e., they are not extra instances in the data set.

3. RESULTS

We fit a total of 23 models to the Assisstments data, using the `glmer` function in the R package `lme4` to fit all models. We compared models in terms of AIC and BIC. Both ranked our models in the same order for this data, so we only report AIC scores in Figure 1.

All models that include R_{ij} (R-PFA and R-only) outperform all existing models by a wide margin. As in previous work, PFA outperforms AFM [6]. Finally, the count of prior successes alone (sPFA) is a better predictor than total opportunities (AFM).

When $b = 1$, R_{ij} is the overall proportion of successes, so R-PFA and R-only use the entire history of performance. Yet, proportion of success (R-PFA and R-only) is more predictive than the number of successful attempts (PFA). For any fixed decay parameter b, R-PFA is better than R-only. The total number of practice opportunities is still informative above and beyond the recent history.

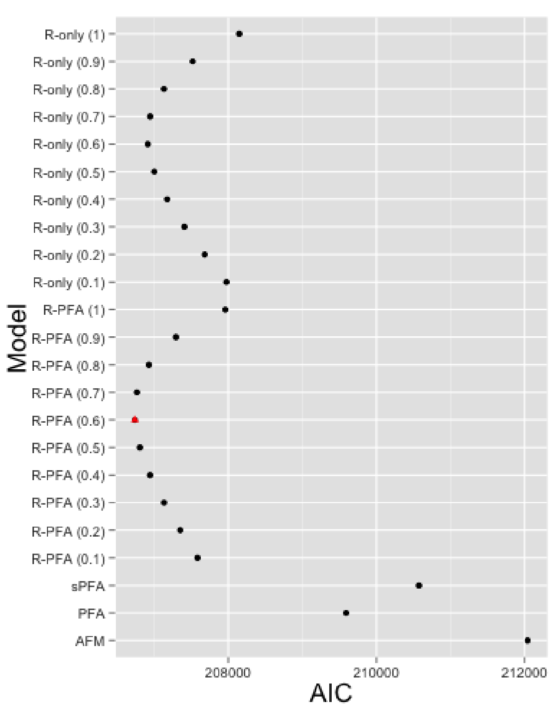

Figure 1: AIC scores for all models (lower is better). The red triangle indicates the lowest AIC.

4. DISCUSSION & CONCLUSIONS

The R-PFA model differs from PFA in two significant ways. First, unless $b = 1$, R-PFA values recent evidence more than older evidence. Second, the ghost attempts reduce the noise of predictions on the first few attempts that a student makes on any particular KC by incorporating the belief that students are unlikely to already know the KC and are unlikely to perform well on a skill on their early attempts. With these two modifications, all variants of R-PFA outperform PFA and other models in terms of predictive accuracy. Ghost attempts and decay weights matter in combination. The ghost attempts necessarily have the greatest influence on practice strings that are relatively short, and there are many such occurrences in our dataset. The ghost attempts reduce the noise that would otherwise be present in R_{ij} for these attempts. The weighting that controls the window of "recent" performance considered is also key. The difference in AIC scores between R-PFA with $b = 0.6$ and $b = 1$ is as great as the difference between sPFA and PFA. However, while R-PFA with $b = 0.6$ performed best on this dataset, that specific value of b may be due to the mastery criterion (a streak between 3-5 trials) in the Assistments software.

There are a number of tunings of R-PFA that we may explore in future work. First, there may be a relationship between the optimal number of ghosts attempts and the decay parameter b. Second, should there be different b values for different knowledge components? Third, should first-attempt hint requests be distinguished from first-attempt incorrects, by incorporating separate proportions for these prior practice outcomes? We hope that R-PFA sees widespread use in the toolset of educational data mining.

5. REFERENCES

[1] R. S. Baker, A. B. Goldstein, and N. T. Heffernan. Detecting learning moment-by-moment. In *IJAIED*, volume 21, pages 5–25, 2011.

[2] H. Cen, K. Koedinger, and B. Junker. Learning factors analysis – a general method for cognitive model evaluation and improvement. In *Proceedings of 8th ITS Conference*, pages 164–175, Berlin / Heidelberg, 2006. Springer.

[3] P. de Boeck and M. Wilson, editors. *Explanatory item response models: a generalized linear and nonlinear approach.* Springer, New York, 2004.

[4] A. Hershkovitz, R. S. Baker, S. M. Gowda, and A. T. Corbett. Predicting future learning better using quantitative analysis of moment-by-moment learning. In *Proceedings 6th EDM Conference*, pages 74–81, 2013.

[5] K. R. Koedinger, A. T. Corbett, and C. Perfetti. The knowledge-learning-instruction framework: Bridging the science-practice chasm to enhance robust student learning. *Cognitive Science*, 36(5):757–798, 2012.

[6] P. I. Pavlik, H. Cen, and K. Koedinger. Performance factors analysis –a new alternative to knowledge tracing. In *Proceedings of AIED*, pages 531–538. IOS Press, 2009.