Pinpointing Learning Moments; A finer grain P(J) model

Adam B. Goldstein[1], Ryan S.J.d. Baker[2], Neil T. Heffernan[1]

[1]Department of Computer Science, Worcester Polytechnic Institute
100 Institute Road, Worcester MA 01609, USA
abgoldstein@gmail.com, nth@wpi.edu
[2]Department of Social Science and Policy Studies, Worcester Polytechnic Institute
100 Institute Road, Worcester MA 01609, USA
rsbaker@wpi.edu

1 Introduction

Educational data mining and knowledge engineering methods have led to increasingly precise models of students' knowledge as they use intelligent tutoring systems. The first stage in this progression was the development of Bayes Nets and Bayesian frameworks that could infer the probability that a student knew a specific skill at a specific time from their pattern of correct responses and non-correct responses (e.g. errors and hint requests) up until that time [cf. 2, 4, 5].

However, while the extensions made in recent years to educational data mining have created the potential for more precise assessment of student knowledge at a specific time, these models do not tell us *when* the knowledge was acquired. Baker, Goldstein, and Heffernan proposed the idea of a model that can infer the probability that a student learned a skill at a specific step during the problem-solving process. This model, P(**J**) for **J**ustLearned, was shown to be a consistent predictor of high eventual $\mathbf{L_n}$ values if a spike in P(**J**) is seen. This ability can potentially allow for engineering intelligent tutoring systems to bias content in a way that can induce these moments of learning. The prior approach achieved a correlation coefficient of 0.446 using only first response features, so we decided to expand the feature set to include subsequent actions. We will discuss the application of P(**J**) to a second ITS, ASSISTments, and we will also compare models that do and do not include subsequent actions.

2 The P(J) Model

The original analysis of P(**J**) used data from 232 students' use of a Cognitive Tutor curriculum for middle school mathematics [3], during the 2002- 2003 school year. These students made 581,785 transactions (either entering an answer or requesting a hint) on 171,987 problem steps covering 253 skills. In [1] it was demonstrated that a model as described above can be created. This model calculates P(**J**ustLearned), P(**J**) for short, which is the probability that a student just learned a skill after a certain problem step. This concept can be expressed in terms of BKT as $P(\mathbf{\sim L_n \wedge T \mid A_{+1+2}})$. For each problem step, [1] used a set of 25 features describing the first action on problem step **N**. These features had in turn been used in prior work to develop automated detectors of off-task behavior [2] and gaming the system. This attempt at a P(**J**) model achieved a correlation coefficient of 0.446 when running Linear Regression in RapidMiner.

3 Refinements to P(J)

After our original model, we hypothesized that we could capture a significant amount of additional variance in fitting the P(**J**) label through features that covered subsequent actions, i.e. tracking what happens following a first response that is wrong or is a help request. We decided to test this theory while also porting the model to data from ASSISTments. This new data set is pulled from 4187 students from New England middle and high schools that were using the system from 2008-2010. This data includes 55 unique skills and 418,513 logged actions. One benefit of using ASSISTments is the fact that it has extensive information about scaffolding problem steps, as well as – importantly – subsequent actions after a student's first attempt at answering. Our new feature set has 67 features, including information about time spent on scaffolding, number of hints used in scaffolding, and so on. This new model achieves a correlation coefficient of 0.429 when running the same Linear Regression in Rapid Miner with 6 fold student-level cross validation, which seems to suggest the model transitioned well to ASSISTments.

In contrast to our expectations, we did not find that subsequent action features gained much variance. When running the same earlier 25 feature model with ASSISTments data we still achieved a correlation coefficient of .398. This seems to indicate that a model to predict P(**J**) can be achieved without the effort of distilling subsequent action features.

Acknowledgements. This research was supported via NSF grant award number DGE0742503, by the Pittsburgh Science of Learning Center, NSF award number SBE-0836012, and by U.S. Department of Education program grant #R305A070440.

4 References

[1] Baker, Ryan S.J.d, Goldstein, Adam B., Heffernan, Neil T. : Detecting the Moment of Learning. *Intelligent Tutoring Systems*, 10, 26-35 (2010).

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

[3] Koedinger, K.R.: Toward evidence for instructional principles: Examples from Cognitive Tutor Math 6. In: *Proceedings of PME-NA XXXIII (the North American Chapter of the International Group for the Psychology of Mathematics Education*) (2002).

[4] Martin, J., VanLehn, K.: Student Assessment Using Bayesian Nets. *International Journal of Human-Computer Studies*, 42, 575-591 (1995).

[5] Shute, V.J.: SMART: Student modeling approach for responsive tutoring. *User Modeling and User-Adapted Interaction*, 5 (1), 1-44 (1995).