

# Tracking Students' Inquiry Paths through Student Transition Analysis

Matt Bachmann<sup>1</sup>, Janice Gobert<sup>1,2</sup> and, Joseph Beck<sup>1,2</sup>

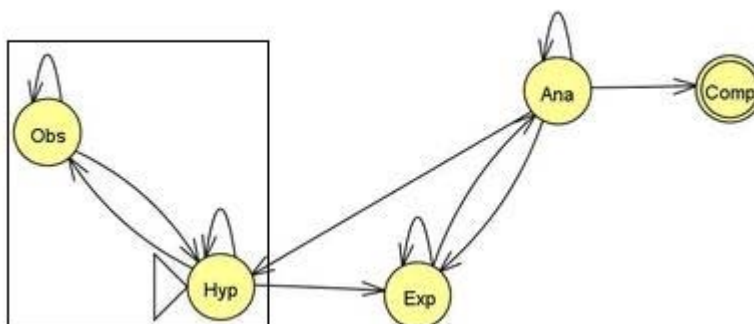
{matt.bachmann, jgobert, joseph.beck}@wpi.edu

<sup>1</sup>Computer Science Department,

<sup>2</sup>Social Sciences & Policy Studies Dept.

Worcester Polytechnic Institute

In the Science Assistentment project ([http://users.wpi.edu/~sci\\_assistments/](http://users.wpi.edu/~sci_assistments/)) we provide students with rich and dynamic opportunities to engage in inquiry while researchers and teachers use performance data as assessment indices of scientific inquiry skills and domain knowledge. Activities in the Science Assistentment Project follow a pedagogical model known as the inquiry cycle, namely, explore, hypothesize, experiment, and analyze. Figure 1 shows this cycle as a graph and the transitions that can be made between states. State “Comp” represents completion of the activity.



**Figure 1. The Inquiry process as modeled by our activity.**

Important to researchers and developers are ways to assess students' inquiry process within microworlds, as well as a rigorous method for providing scaffolding to students who struggle with inquiry during the activity. The method presented here utilizes a Markov chain to track a student's path through the various inquiry activities.

The data were collected from 148 eighth grade students, ranging in age from 12-14 years, from a public middle school in Central Massachusetts. Students engaged in inquiry using our learning environment including a microworld simulating the phase changes of water and a series of inquiry support tools for making a hypothesis, collecting data, and analyzing data.

Based on the logging features of these microworlds, we track the path students take through the activity by counting the transitions students make and create a Markov model of their inquiry process (Figure 1). This potentially affords researchers a way of grouping students based on their path through the activity. The rationale for this is that some paths are more effective representing systematic inquiry [1] and, in turn, lead to positive learning gains while other inquiry paths do not. Using this Markov model, researchers hope to eventually be able to develop scaffolds for students whose path corresponds to less effective inquiry. However, in order to do this we need have models that can diagnose a students' inquiry path. The current work seeks to develop such a student model.

The first attempt at breaking up different groups of students used the scores students obtained on the multiple choice pretest: those in the top third on the pre-test, and those in the bottom third. The middle third was not used for this analysis. Based on these two groups of students, we generated two transition models. To validate these models we computed the log likelihood of each model given each student. A paired samples t-test on the log likelihoods of the intended models (the model generated from the group which the student belonged) versus the unintended models (a model generated from students from a different group) found the models to be distinct ( $t(98) = 3.385, p < .001$ ). Using a sign test (log likelihood of intended model - log likelihood of unintended model) for each student, we found that the intended model had a higher log likelihood 68% of the time. Although this disaggregation was somewhat successful, 68% accuracy in terms of classification is not a great improvement over a baseline of 50% generated by random guessing.

For our next attempt, rather than picking some measure and using that to disaggregate the models, we looked at the models themselves and attempted to disaggregate them by applying a K-means cluster analysis using Weka (version 3.6.2), which generated two models each of which was based on a cluster. The major difference between the two models is that students in model 1 tended to return to the observe state of the inquiry cycle (see highlighted portion of Figure 1).

The next steps going forward are to try to use our models along with other learner characteristics we have about our participants to predict what cluster the students fall into. We also need to find if the clusters fit into categories of high and low performance in the activity.

## References

- [1] Buckley, B.C, Gobert, J.D., Horwitz, P. (2006) Using Log Files To Track Students' Model-based Inquiry. *Proceedings of the Seventh International Conference of the Learning Sciences (ICLS)*, Mahwah, NJ: Erlbaum, p.57-63.