A Computational Model of Accelerated Future Learning through Feature Recognition

Nan Li, William W. Cohen, and Kenneth R. Koedinger

School of Computer Science
Carnegie Mellon University
5000 Forbes Ave., Pittsburgh PA 15213 USA
{nli1,wcohen,koedinger}@cs.cmu.edu

Abstract. Accelerated future learning, in which learning proceeds more effectively and more rapidly because of prior learning, is considered to be one of the most interesting measures of robust learning. A growing body of studies have demonstrated that some instructional treatments lead to accelerated future learning. However, little study has focused on under- standing the learning mechanisms that yield accelerated future learning. In this paper, we present a computational model that demonstrates accelerated future learning through the use of machine learning techniques for feature recognition. In order to understand the behavior of the proposed model, we conducted a controlled simulation study with four alternative versions of the model to investigate how both better prior knowledge learning and better learning strategies might independently yield accelerated future learning. We measured the learning outcomes of the models by rate of learning and the fit to the pattern of errors made by real students. We found out that both stronger prior knowledge and a better learning strategy can speed up the learning process. Some model variations generate human-like error patterns, but others learn to avoid errors more quickly than students.

1 Motivation and Algorithm

Perhaps one of the most interesting measures of robust learning is accelerated future learning. A growing number of studies have experimentally demonstrated that some instructional treatments lead to accelerated future learning. These treatments (and associated studies) include inventing for future learning [1], self-explanation [2], and feature prerequisite drill [3]. While results are starting to accumulate, we have little by way of precise understanding of the learning mechanisms that yield these results. A computational model of accelerated future learning that fits student learning data would be a significant achievement in theoretical integration within the learning sciences, and reveal insights on improving current education technologies.

Previous work [4] showed that one of the key factors that differentiates experts and novices is that experts view the world in terms of deep functional features, while novices see in terms of shallow perceptual features. In this paper, we propose a novel approach to modeling accelerated future learning through the use of machine learning techniques to acquire deep features. We assume that the

input of the system is a set of feature recognition records. Each record consists of an original problem (e.g. an expression, $-3x$), and the feature recognized from the problem (e.g. the coefficient in the problem, -3 in $-3x$). The objective of this work is to construct a computational model to learn feature recognition.

After careful examination of the problem, we find out that the feature recognition problem closely connects to the probabilistic context free grammar (PCFG) induction problem, where knowledge is represented by grammar rules, and the learning process is similar to grammar induction. Therefore, we extended a grammar induction algorithm proposed by Li et al. [5], since it acquires PCFG from observation sequences without any prior structural knowledge. Details about this learning algorithm are described in [5].

To support feature learning, after acquiring the grammar with Li et al.'s algorithm, our learning system finds the intermediate symbol that corresponds to the feature most frequently in the parse trees of the training examples, and identifies it as the target feature. To understand how prior knowledge and learning strategy could affect learnng outcomes, we extended the learning algorithm in two directions. First, we designed a transfer learning mechanism that biases the probabilities of rules in future tasks toward the probabilities associated with previous tasks. The learner records the number of times each grammar rule appeared in a parse tree from previous tasks, and updates the rule probability in a new task by adding the previous applied rule frequency to the training problems. Second, we extended our learning mechanism to making use of a "semantic non-terminal constraint" embedded in training data during learning. More specifically, the learner forces all the feature subsequences to correspond to one non-terminal symbol.

2   Empirical Study

We carried out a controlled simulation study in algebra to test 1) whether stronger prior knowledge and better learning strategies could yield accelerated future learning, 2) if so, how prior knowledge and learning strategies affect the learning outcome. There were 2-by-2 (4) alternative versions of the proposed learning model in the study: L00, the original learner without transfer learning and the non-terminal constraint; L01, the learner with the non-terminal constraint but without transfer learning; L10, the learner with transfer learning but without the non-terminal constraint; L11, the learner with both tranfer learning and the non-terminal constraint.

We designed three curricula in the study. Three tasks were used across the three curricula with increasing complexities. The three curricula are 1) task one, then task two; 2) task two, then task three; 3) task one, then task two, then task three. In all but the last task, each learner was given 10 training problems. For the last task, each learner was given one to five training records. Under each training condition, both systems were tested on 100 expressions in the same form of the training data in the last task. For each testing record, we compared the feature recognized by the oracle schemas with that recognized by the acquired schemas, and evaluated the correctness of output.

We also compared the errors made by the learning system with common errors made by real students in curriculum one. We inspected the common errors
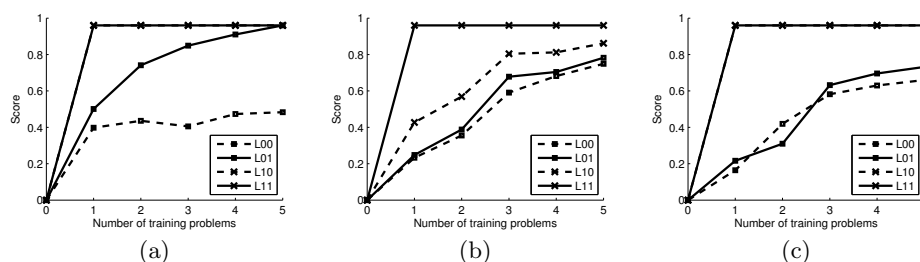
Fig. 1. Learning curves for four learners in curriculum (a) from task one to task two. (b) from task two to task three. (c) from task one to task two to task three.

made by real students in task two from a study of 71 high school students used Carnegie Learning Algebra I Tutor, and noticed that mishandling of negative coefficients (e.g., marking 3 instead of -3 as the coefficient of $-3x$) is the most common error. The learning system was asked to recognize the coefficients of the 100 given expressions, and was evaluated based on the match of the errors made by the learner and the most common error made by real students.

As shown in Figure 1(a), the result suggests that with transfer learning, learners are able to acquire knowledge quicker than those without transfer learning. Comparing the base learner, L00, and the learner with non-terminal constraint, L01, we can see that a better learning strategy yields a steeper learning curve. We can also see that in all three curricula, the transfer learner, L10, always outperforms the learner with semantic non-terminal constraint, L01. Similar results were also observed with curriculum two and curriculum three. This suggests that prior knowledge is more effective in accelerating future learning than better learning strategies. In the error matching study, we see that after being trained with one to five problems, L00 generated the most common error in testing. Besides that all other incorect answers are due to the incapability of identifying a coefficient from the problem.

References

1. Bransford, J.D., Schwartz, D.L.: Rethinking transfer: A simple proposal with multiple implications. Review of Research in Education **24** (1999) 61–100
2. Hausmann, R.G., VanLehn, K.: Explaining self-explaining: A contrast between content and generation. Artificial intelligence in education: Building technology rich learning contexts that work **158** (2007) 417–424
3. Pavlik, Jr., P., Bolster, T., Wu, S.M., Koedinger, K., Macwhinney, B.: Using optimally selected drill practice to train basic facts. In: Proceedings of the 9th international conference on Intelligent Tutoring Systems, Berlin, Heidelberg (2008) 593–602
4. Chi, M.T.H., Feltovich, P.J., Glaser, R.: Categorization and representation of physics problems by experts and novices. Cognitive Science **5**(2) (1981) 121–152
5. Li, N., Kambhampati, S., Yoon, S.: Learning probabilistic hierarchical task networks to capture user preferences. In: Proceedings of the 21th International Joint Conference on Artificial Intelligence, Pasadena, CA, USA (2009)