Leveraging Semi-Supervised Learning to Predict Student Problem-Solving Performance in Narrative-Centered Learning Environments

Wookhee Min, Bradford Mott, Jonathan Rowe, and James Lester

North Carolina State University, Raleigh, North Carolina, USA
{wmin, bwmott, jprowe, lester}@ncsu.edu

Abstract. This paper presents a semi-supervised machine-learning approach to predicting whether students will be successful in solving problem-solving tasks within narrative-centered learning environments. Results suggest the approach often outperforms standard supervised learning methods.

Keywords: Narrative-centered learning environments, game-based learning environments, semi-supervised learning.

1 Introduction

Recent years have witnessed growing interest in narrative-centered learning environments, which tightly integrate interactive narratives, digital games, and the adaptive pedagogy of intelligent tutoring systems to generate highly engaging interactive story experiences for personalized learning [1]. Because students have considerable autonomy in these open-ended environments, it is possible for students to unintentionally spend time on problem-solving tasks for which they already have mastery, and inadvertently skip problem-solving tasks where they have gaps in knowledge. This paper introduces a data-driven method for predicting whether students will successfully complete problem-solving tasks based on their prior performance. We leverage self-training semi-supervised learning as a framework for predicting problem-solving task success [2]. We compare this framework to naïve Bayes (NB) and support vector machine-based (SVM) classifiers. Results suggest that self-training often provides the most accurate predictions. The resulting models show significant promise for supporting pedagogical planning in narrative-centered learning environments.

2 Results and Discussion

To evaluate the self-training semi-supervised learning approach to predicting problem-solving performance, we analyze student data from a classroom deployment of

CRYSTAL ISLAND [1]. During game play, students progressed in solving the problem scenario by completing concept matrices based on informational texts they encountered in the game. In this work, we analyze student data from 10 frequently attempted concept matrices (on average, 565 students attempted them).

Prediction accuracy rates are compared across self-training semi-supervised learning, supervised learning, and a baseline using the majority label. The results fall into two major categories: (1) the self-training method outperforms the corresponding supervised learning technique and baseline, and (2) the baseline performs better than both self-training and supervised learning. In the first category, 3 out of the 10 classifications show that self-training using NBs outperforms both the other two approaches, and 4 out of the 10 classifications show that self-training using SVMs outperforms the other approaches. Table 1 describes pairwise comparisons using one-way repeated measures ANOVA for NBs ($F(1.13, 32.64) = 74.91$, $p < 0.001$) and SVMs ($F(1.01, 39.55) = 34.98$, $p < 0.001$) for the classifications in this category. Statistical significance is measured using least significant difference post-hoc tests. The second category of observations in which the baseline (M=87.39) performs better than both supervised learning (M=84.12) and self-training (M=84.38) consists of relatively easy problem-solving tasks in which 87.39% of students successfully solved the tasks.

Table 1. Average Model Accuracy on Predicting Success of Problem-Solving Tasks in First Category. (statistical significance over * baseline and § supervised learning)

Approach	Naïve Bayes	Support Vector Machine
Baseline	68.107	74.880
Supervised Learning	71.724^{*}	83.338^{*}
Self-Training	$73.182^{*\S}$	$83.788^{*\S}$

We have proposed an approach to predicting problem-solving performance leveraging semi-supervised learning. Results suggest that the self-training semi-supervised learning method can improve predictive models' accuracy over standard supervised learning techniques, and thus support adaptive pedagogical planning in narrative-centered learning environments.

Acknowledgements. This research was supported by the National Science Foundation under Grant IIS-1138497. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors.

3 References

1. Min, W., Rowe, J., Mott, B., Lester, J.: Personalizing Embedded Assessment Sequences in Narrative-Centered Learning Environments: A Collaborative Filtering Approach. In: 16th International Conference on Artificial Intelligence in Education, pp. 369–378 (2013)
2. Zhu, X., Goldberg, A.: Introduction to semi-supervised learning. Synthesis lectures on artificial intelligence and machine learning. 3(1), 1–130 (2009)