Predicting Preschool Mathematics Performance of Children with a Socially Assistive Robot Tutor

Caitlyn Clabaugh*, Konstantinos Tsiakas† and Maja Matarić*
* Viterbi School of Engineering
University of Southern California
{clabaugh, mataric}@usc.edu
† Department of Computer Science and Engineering
The University of Texas at Arlington
konstantinos.tsiakas@mavs.uta.edu

***Abstract*—This paper considers the problem of modeling the performance of preschool children on mathematics learning activities towards a personalized intelligent tutoring system. It presents a formalism for preschool number concepts and proposes a model for predicting the ability of preschool children to complete number concept activities. The approach is validated on a dataset of preschool child interactions with a socially assistive robot tutor for number concepts. Our results demonstrate the effectiveness of two factors in improving performance prediction: domain knowledge about early childhood mathematics and relative clustering of child participants by performance. We discuss the limitations and design considerations in our approach relative to personalization.**

I. Introduction

In this paper, we consider the problem of modeling the ability of a particular child learner with respect to some set of early mathematics learning tasks. We situate this problem in the context of a socially assistive robot tutor for preschool mathematics.

Every learner is unique, and the benefits of one-on-one tutoring are well-supported by the literature; for example, Bloom's seminal work reports that the average tutored student performed two standard deviations above those who learned via conventional methods [1]. Replication of this phenomenon can be seen in early Human-Computer Interaction (HCI) research on computer-based, intelligent tutoring systems (ITS) [2], followed by emerging work on personalized robotic tutors in the field of Human-Robot Interaction (HRI) [3], [4].

However, personalization is a complex and multifaceted computational problem that requires the tutoring agent and/or system to simultaneously and continually assess, adapt, and leverage a model of the learner's ability and needs [5]. The state of the learner is hidden and dynamic, and the signals are often noisy. Thus, models of a human learner are typically scoped to specific educational goals and exploit domain knowledge about those goals.

Our research considers modeling child learners' performance on preschool mathematics activities, incorporating early education learning theories into our system design and modeling processes, using a *socially assistive robot* (SAR) [6] tutor for *number concepts* [7]. The scope of this paper is limited to the *user modeling* component of personalization, constructing predictive models of individual learners' abilities given some history of interaction. This work leverages a literature-based formalism of number concepts learning to better predict children's abilities to complete SAR-guided mathematics exercises. Our methodology is applied to and evaluated on a dataset of preschool child-robot interactions in a general education classroom. The work presented in this paper is an extended analysis of previous work described in Clabaugh et al. [8], [9].

Fig. 1: A child participant interacting with the SAR tutor to practice number concepts. The system features seven touch-based, space-themed games with interactive feedback from an Aldebaran NAO robot.

II. Related Work

In this section, we situate our work within the research paradigm of user modeling for personalized tutoring in HCI and HRI, ground its methodology in popular theories of early childhood education and mathematics, and offer this paper's contributions as extensions of our previous work [8], [9].

A. Intelligent Tutoring Systems

User modeling is an established research paradigm in HCI (see [10] for a review). The physical embodiment of an intelli-

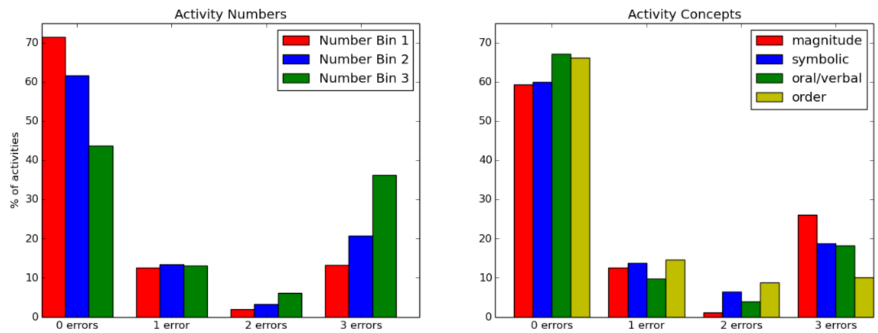

Fig. 2: Percentage of errors based on activities. Left: activity number n in number bins. Right: activity concepts c.

gent tutor has been shown to increase cognitive learning gains [3] and compliance [11]. Additionally, the intrinsic agency of a socially assistive robot tutor lends itself to learning through play [12], a method important to socio-constructivist theories of early childhood learning [13]. Thus, there is motivation to extend ITS user modeling to HRI for early childhood education.

The research presented in this paper deals with the user modeling component of personalization, applying successful methods in ITS (see [14] for a review) to HRI and early childhood mathematics. Our work emphasizes domain-specific scaffolding to mitigate the known computational complexities of user modeling [15].

B. Early Childhood Mathematics

In this paper, we explore modeling performance relative to preschool mathematics, formally called *number concepts*. Number concepts are the building blocks of early childhood mathematics, and include each number's cardinality, nominal representation, and ordinal relations to other numbers [16]. Early childhood mathematics has a strong correlation to children's future educational success [17]. Computational personalization is an approach that may help to satisfy this early childhood educational need. Our analysis uses a child-robot interaction dataset, described in Section III, that inherently includes the engagement potential of robots as social peer tutors [18].

C. Formalizing Number Concepts

As discussed above, many intelligent tutoring systems incorporate domain-specific knowledge to better predict learners' performance. Next, we present a formalism for number concepts learning based on domain knowledge from early childhood mathematics literature.

According to the Transitional Kindergarten Math Pacing Guide [7], we first limit number concepts to include numbers 1 through 9, $N = [1, 9]$. The various concepts C being modeled include: a number's cardinality or *magnitude* M, ordinal relationship to other numbers or *order* O, *symbolic* representation S, and oral or *verbal* representation V.

We thus formalize a learning activity to be some pairing of a number n and concept(s) $c \subset C$ about that number. It is often the case that a learning activity about number concepts contains multiple numbers and/or concepts. We denote an activity α as follows:

$$\alpha = \{n, c \subset C\}$$

Our formalism is based on the definition of number concepts according to early childhood education standards [7], using the standards as domain knowledge to facilitate user modeling for number concepts learning. Our formalism is presented in more detail in Clabaugh et al. [9].

In this work, we further group activities according to the theory of "chunking" in early childhood development [19]. We chunk $n \in [1, 9]$ into three *number bins*: (1) $n \in [1, 2]$, (2) $n \in [3, 4]$, and (3) $n \in [6, 9]$. This chunk-based binning convention is used throughout this work. Figure 2 shows an example.

III. CHILD-ROBOT TUTOR INTERACTION DATASET

In this work, we use features of robot-child interactions to predict performance on randomized sets of number concepts activities. Our approach is applied to and evaluated on a child-robot interaction dataset of 44 preschool participants.

The dataset was collected from a user study with a fully autonomous socially assistive robot (SAR) that acted as a social peer tutor, guiding preschool aged child participants through number concepts activities on a touch tablet computer in a preschool classroom. The system and user study are both described in detail in Clabaugh et al. [8]. For replicability, The SAR number concepts games and the preschool learning styles inventory questionnaire are openly available to researchers.[1].

IV. MODELING NUMBER CONCEPTS LEARNING

We formulate the performance prediction problem as a binary classification task. Our goal is to predict the performance class PC_x, $x \in [0, 1]$, where PC_0 predicts an incomplete activity and PC_1 predicts completion of an activity.

Two important characteristics of the collected dataset inform our approach. First, the classes are imbalanced; completion rate (79%) is much higher than non-completion rate (21%),

[1]For inquiries, please contact contributing author Caitlyn Clabaugh at clabaugh@usc.edu

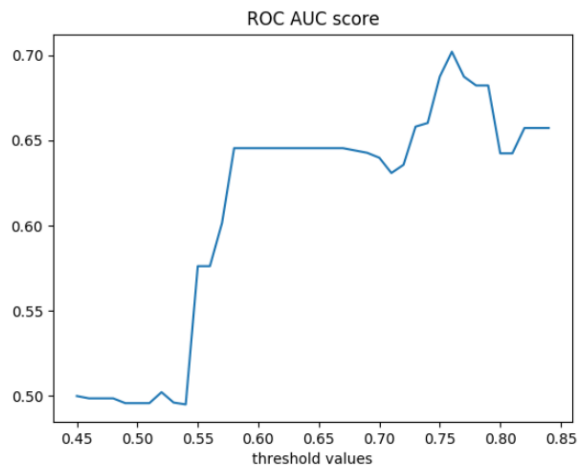

Fig. 3: The ROC AUC scores for different thresholds $t \in T$. The best threshold is the one that maximizes the ROC AUC score, based on the training data.

which we consider typical in properly designed tutoring interactions because most tasks should be within reach of the learner [20]. Second, the data are noisy in terms of child performance across numbers and concepts; this is a common challenge in many user modeling tasks [5].

In the following sections, we describe an approach that takes into consideration domain knowledge about number concepts learning and children's performance relative to each other. We compare the experimental results of our approach relative to standard Support Vector Machines (SVM) and Random Forest (RF) classifiers.

A. *Probabilistic Approach*

Since deterministic classifiers (e.g., SVM, RF) may not be appropriate for this task, considering the dataset characteristics mentioned above, we followed a probabilistic approach to build a classifier that predicts the probability of completion by estimating $P(C_1|n,c)$. In order to train this model, we applied a Maximum Likelihood Estimator (MLE) on the observed data, followed by a Support Vector Regression (SVR) model to learn the probabilities for all possible activities α.

More specifically, the MLE learns the probabilities based on the observed activities:

$$P(PC_1|n,c) = \frac{P(PC_1,n,c)}{P(n,c)}$$

In order to learn the probabilities for each activity (all combinations of numbers and concepts), we employ an SVR that estimates $P(PC_1|n,c)\ \forall\ n,c \subset C$. Given the probability for a specific activity, we need to determine the threshold upon which we decide the output class. This threshold is estimated based on the training data as:

$$T^* = \underset{t \in T}{\operatorname{argmax}}\, ROC\_AUC(t)$$

Figure 3 shows the ROC AUC scores for different threshold values for our specific training set. We use ROC AUC score because of the imbalanced nature of our classes; ROC curves are insensitive to class balance, whereas classification accuracy is informative for balanced classes.

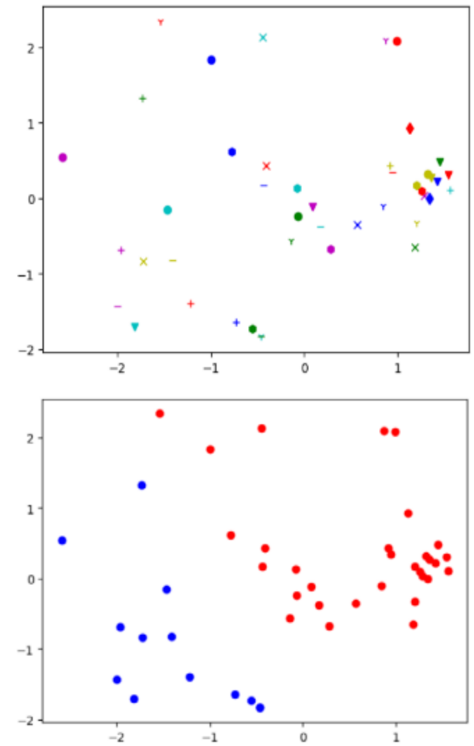

Fig. 4: 2D visualization of performance-based clustering. Top figure shows MDS applied to child participants' performance with respect to $\alpha = \{n, c \subset C\}$. Bottom figure shows k-means clustering applied to MDS of child participants.

B. *Performance-Based Clustering*

In order to gain further insight into the performance distribution of child participants, we project features of their performance into a 2D visualization using multidimensional scaling (MDS). Each child is defined by a performance matrix PM, where PM_{ij} is the average activity completion rate for number bin i and concept j. As seen in Figure 4, we apply MDS to project each matrix into a 2D visualization with each point corresponds to a specific individual. We then apply k-means clustering (K = 2) to the resulting projection such that two clusters of children are defined with respect to their performances. Towards the ultimate goal of personalization, we incorporate this clustering into our model training and evaluation described in the following section.

V. Evaluation of Models

To evaluate our approach, we compare it with that of an SVM (linear kernel) and RF (100 estimators). Specifically, we used the built-in method of the Python `sklearn.svm.SVC` library to estimate the probabilities $P(PC_x|n,c)$ for PC_0 and PC_1 based on the SVM model training (probSVM). Based on these probabilities, the predicted class is the one with

	RF	SVM	probSVM	MLE+SVR	Voting	Clustered
AUC ROC	0.63	0.64	0.52	0.64	0.65	**0.67**
FNR	0.38	0.37	0.06	0.47	0.41	0.29
FPR	0.36	0.34	0.81	**0.27**	0.29	0.35

TABLE I: Experimental results of performance prediction models. AUC ROC indicates the area under the ROC curve, FNR indicates the false negative rate, and FPR indicates the false positive rate of the trained classifier.

the largest probability. We also employ a voting classifier to combine the decisions of the other classifiers, as a sort of ensemble method.

Lastly, we use our MDS and K-Means approach to train all classifiers, using only the data from one cluster. We used the larger cluster (red in Figure 4) and created a smaller dataset, including the corresponding participant data. The results are shown in Table I.

We employ leave-one-out validation to ensure generalization across individual participants. For each of the N ($N = 44$) users, we include the $N-1$ in the training set and test the model on the remaining individual. The final score is the average score over all participants.

To compare these different classifiers, we consider their ROC AUC scores and the False Positive (FPR) and False Negative Rates (FNR). The results, including our approach (MLE+SVR), are shown in Table I.

VI. Discussion

The main observation from this work is that there is a slight improvement in the scores using our proposed method and the voting classifier. However, we also observe the challenge of minimizing both FPR and FNR.

Relative to our end-application requirements, a high FPR means that failure to complete an activity would not be predicted by the robot, resulting in absence of intervention when needed. A high FNR would increase the number of robot interventions or support, including activities where intervention is not needed, which prioritized FPR minimization. We observe that our method minimizes FPR, compared to SVM and RF, and overall, the voting classifier maximizes the ROC AUC score.

Lastly, we observe an improvement in ROC AUC score and a better balancing of FPR and FNR, while FPR increases. We take as an indication that personalized prediction models would be more accurate, and thus, result in more efficient interactions.

The classification problem considered is constrained to a binary outcome of performance and our approach is evaluated on a dataset specific to preschool children. However, the results of this work demonstrate the importance of leveraging both domain knowledge and similarities among learners to improve performance prediction.

Acknowledgment

This work was supported by an NSF Expedition in Computing IIS-1139148.

References

[1] B. S. Bloom, "The 2 sigma problem: The search for methods of group instruction as effective as one-to-one tutoring," *Educational researcher*, pp. 4–16, 1984.

[2] A. Corbett, "Cognitive computer tutors: Solving the two-sigma problem," in *User Modeling 2001*. Springer, 2001, pp. 137–147.

[3] D. Leyzberg, S. Spaulding, M. Toneva, and B. Scassellati, "The physical presence of a robot tutor increases cognitive learning gains," in *34th Annual Conference of the Cognitive Science Society (Cog. Sci.12)*, 2010.

[4] J. Greczek, E. Kaszubski, A. Atrash, and M. Matarić, "Graded cueing feedback in robot-mediated imitation practice for children with autism spectrum disorders," 2010.

[5] G. I. Webb, M. J. Pazzani, and D. Billsus, "Machine learning for user modeling," *User modeling and user-adapted interaction*, vol. 11, no. 1-2, pp. 19–29, 2001.

[6] M. J. Mataric, "Socially assistive robotics," *Science*, vol. 3, no. 6, pp. 233–242, 1999.

[7] *Transitional Kindergarten Math Pacing Guide*, Paramount Unified School District, 2013.

[8] C. Clabaugh, G. Ragusa, F. Sha, and M. Matari'c, "Designing a socially assistive robot for personalized number concepts learning in preschool children," in *Fifth International Conference on Development and Learning and on Epigenetic Robotics (ICDL-EPIROB*, Providence, RI, August 2015, pp. 314–319.

[9] C. Clabaugh, F. Sha, G. Ragusa, and M. Matari'c, "Towards a personalized model of number concepts learning in preschool children," in *ICRA Workshop on Machine Learning for Social Robotics*, Seattle, Washington, May 2015.

[10] J. R. Anderson, C. F. Boyle, A. T. Corbett, and M. W. Lewis, "Cognitive modeling and intelligent tutoring," *Artificial intelligence*, vol. 42, no. 1, pp. 7–49, 1990.

[11] W. A. Bainbridge, J. Hart, E. S. Kim, and B. Scassellati, "The effect of presence on human-robot interaction," in *Robot and Human Interactive Communication, 2008. RO-MAN 2008. The 17th IEEE International Symposium on*. IEEE, 2008, pp. 701–706.

[12] E. Short, K. Swift-Spong, J. Greczek, A. Ramachandran, A. Litoiu, E. C. Grigore, D. Feil-Seifer, S. Shuster, J. J. Lee, S. Huang *et al.*, "How to train your dragonbot: Socially assistive robots for teaching children about nutrition through play."

[13] D. G. Singer, R. M. Golinkoff, and K. Hirsh-Pasek, *Play= Learning: How play motivates and enhances children's cognitive and social-emotional growth*. Oxford University Press, 2006.

[14] E. Wenger, *Artificial intelligence and tutoring systems: computational and cognitive approaches to the communication of knowledge*. Morgan Kaufmann, 2014.

[15] P. Langley, "Machine learning for adaptive user interfaces," in *KI-97: Advances in artificial intelligence*. Springer, 1997, pp. 53–62.

[16] M. J. Behr, R. Lesh, T. R. Post, and E. A. Silver, "Rational number concepts," *Acquisition of mathematics concepts and processes*, pp. 91–126, 1983.

[17] F. A. Campbell, C. T. Ramey, E. Pungello, J. Sparling, and S. Miller-Johnson, "Early childhood education: Young adult outcomes from the abecedarian project," *Applied Developmental Science*, vol. 6, no. 1, pp. 42–57, 2002.

[18] T. Kanda, M. Shimada, and S. Koizumi, "Children learning with a social robot," in *Proceedings of the seventh annual ACM/IEEE international conference on Human-Robot Interaction*. ACM, 2012, pp. 351–358.

[19] D. H. Clements and J. Sarama, "Rethinking early mathematics: What is research-based curriculum for young children?" in *Reconceptualizing early mathematics learning*. Springer, 2013, pp. 121–147.

[20] M. A. Guadagnoli and T. D. Lee, "Challenge point: a framework for conceptualizing the effects of various practice conditions in motor learning," *Journal of motor behavior*, vol. 36, no. 2, pp. 212–224, 2004.