Elaine Short, David Feil-Seifer, and Maja Mataric. "A Comparison of Machine Learning Techniques for Modeling Human-Robot Interaction with Children with Autism". To appear in Proceedings of the International Conference on Human-Robot Interaction (HRI), Lausanne, Switzerland, Mar 2011.

A Comparison of Machine Learning Techniques for Modeling Human-Robot Interaction with Children with Autism

Elaine Short
Univ. of Southern California
Dept. of Computer Science
Los Angeles, California, USA
elaine.g.short@usc.edu

David Feil-Seifer
Univ. of Southern California
Dept. of Computer Science
Los Angeles, California, USA
dfseifer@usc.edu

Maja Matarić
Univ. of Southern California
Dept. of Computer Science
Los Angeles, California, USA
mataric@usc.edu

ABSTRACT

Several machine learning techniques are used to model the behavior of children with autism interacting with a humanoid robot, comparing a static model to a dynamic model using hand-coded features. Good accuracy (over 80%) is achieved in predicting child vocalizations; directions for future approaches to modeling the behavior of children with autism are suggested.

Categories and Subject Descriptors

H.1.2 [**Models and Principles**]: User/Machine System; I.2.9 [**Robotics**]: Miscellaneous

General Terms

Performance, Design, Experimentation, Human Factors

Keywords

Human-robot interaction, machine learning, autism

1. INTRODUCTION

The use of robotic systems is a promising technological possibility for enhancing therapy for children with autism, a common and often debilitating developmental disorder affecting between one in 80 and one in 240 children in the United States [6]. Anecdotal evidence and case studies suggest that not only are robots highly salient to children with autism, but that those children may exhibit social behaviors with robots that they do not otherwise use (e.g., [7]). A number of research groups (including our own) have used robots with children with autism (e.g., [7], [3], and [2]). On the machine learning side, modeling the behavior of children with autism has mainly focused on diagnosis; machine learning techniques have been used to discriminate between children with autism and typically developing children such as in [5] and [9]. Other work has attempted to model behavior in children with autism, but has either been focused on nonsocial behavior ([1]) or has been limited in the generalizations that could be made due to the heterogeneity of the population ([8]). The approach of this work is unique in the use of multimodal features to model social behaviors in children with autism. We use a combination of audio and video features in order to identify one particular social behavior: child vocalizations. We use two different machine learning techniques to model the interaction in order to predict vocalizations.

HRI'11, March 6–9, 2011, Lausanne, Switzerland.
ACM 978-1-4503-0561-7/11/03.

2. EXPERIMENTAL DESIGN

The dataset comes from a study comparing children's interactions with a robot behaving in a way that is contingent on the child's behavior, a robot that behaves randomly, and a non-robotic toy. A description of the system used appears in [2]. There are three primary experiment conditions, but for this preliminary work we use data from sessions for six children with autism interacting with the humanoid robot where the robot's behavior is contingent on the child's behavior. At this time, these data are annotated by a single coder (the author). There are 18 features, with 44 total possible feature-value pairs, including such features as where the child was standing and whether the child was touching the robot, the wall, or the parent. Additionally, the PrAAt audio analysis software was used to extract pitch and intensity features from the audio.

Two machine learning algorithms are used in the analysis of the data. Conditional Random Fields (CRF) are used because their dynamic nature and ability to both classify and segment data are well-suited to the time-series data generated by the experiment (using code by Kevin Murphy [4]). As a comparison, we also examine the performance of decision trees (generated using the C4.5 algorithm in the Weka toolkit). One social behavior of particular interest to us is child vocalizations, since that behavior is used in the experiment as one that receives the "reward" behavior from the robot (blowing bubbles). Thus we focus on the recognition of vocalizations, with the following hypotheses:

H1: CRF will outperform decision trees for recognition of child vocalizations.

H2: Recognition using the full multimodal feature set will outperform recognition using only audio features.

3. RESULTS

Additional time-shifted features were added to the feature set five, ten, and fifteen frames in the past (up to half of a second). This results in a total feature set size of 241 features. The learning algorithms were evaluated using leave-one-out validation, first on the full set of features, then on the set of audio features only. CRF was additionally tested on a subset of 64 optimal features chosen using information gain. Additional analysis was performed with k-means clustering, but is beyond the scope of this abstract.

Because of the limited numbers of children on whose data we could train the models, there was limited statistical significance in the results. The only statistically significant pairwise difference in performance was for the F1 values in the audio only case, with $p < 0.05$.

For the full set of data, the CRF yielded a mean F1-value of 0.0889, with a variance of 0.0156, and an error rate of 26.35% with a variance of 1.62%. The decision tree gave an F1 value of 0.1181 (variance: 0.0017) and an error rate of 28.19% (var: 1.74%). For the audio-only data, the CRF's mean F1 value was 0.1784 (var: 0.0035) and had an error rate of 21.48% (var: 3.22%). The decision tree had a mean F1 value of 0.0714 (var: 0.0021) and an error rate of 20.02%(var: 2.51%). Finally, for the set of best features, the CRF had a mean F1 value of 0.2281 (var:0.0700) and an error rate of 19.81% (var:1.61%).

4. DISCUSSION

Hypothesis 1: CRF outperforms decision trees: The outcome of this hypothesis depends on the set of features used. Although the conditional random field and the decision tree seem to perform comparably in terms of error rates, when we look at the F1 value (the harmonic mean of precision and recall), we see that the decision tree outperforms the conditional random field in the set of all data, while the conditional random field outperforms the decision tree on audio features only, as well as on the set of best features. Furthermore, CRF on audio features only outperforms the decision tree on the set of all features. This is likely because the decision tree is a pruned decision tree, which prevents over-fitting on the data, and desensitizes it to noise. On the full set of data, CRF (which does not reduce the feature set in any way) is hampered by the noise as well as by over-fitting. Our hypothesis is tentatively confirmed, with the caveat that, on a large feature set, it will be necessary to somehow reduce the size of the feature set.

Hypothesis 2: Multimodal feature set outperforms audio-only feature set: The data do not support this hypothesis for CRF in the comparison between the audio and full feature sets. However, as discussed above, this is likely because an overabundance of features caused overfitting and an increase in the model's sensitivity to noise. When we reduced the feature set size to the 64 best features, we achieved performance above that of the audio-only feature set, confirming that multimodal information is useful for predicting child vocalizations.

5. CONCLUSIONS

The population of individuals with autism spectrum disorders (ASD) is exceedingly heterogenous when it comes to social behavior–hence the difficulty of both diagnosis and treatment. The primary contribution of this work is the successful application of machine learning techniques to model vocalization behavior in children with ASD. We have used machine learning to model one social behavior with relatively high accuracy. We have also demonstrated that conditional random fields outperform decision trees with this particular dataset, suggesting that future work should focus on the use of dynamic rather than static classifiers in order to capture temporal structure in the data of this type.

6. ACKNOWLEDGMENTS

We gratefully acknowledge the resources provided by the Autism Genetic Resource Exchange (AGRE) Consortium and the participating AGRE families. The Autism Genetic Resource Exchange is a program of Autism Speaks and is supported, in part, by grant 1U24MH081810 from the National Institute of Mental Health to Prof. Clara M. Lajonchere (PI). This work is supported by the Nancy Lurie Marks Family Foundation, the LA Basin Clinical and Translational Science Institute, the Dan Marino Foundation, and the National Science Foundation (CNS-0709296, IIS-0803565, and Graduate Research Fellowship Program).

7. REFERENCES

[1] F. Albinali, M. S. Goodwin, and S. S. Intille. Recognizing stereotypical motor movements in the laboratory and classroom: a case study with children on the autism spectrum. In *Proc. of the 11th Int. Conf. on Ubiquitous Computing*, Ubicomp '09, pages 71–80, New York, NY, USA, 2009. ACM.

[2] D. Feil-Seifer and M. Matarić. Toward socially assistive robotics for augmenting interventions for children with autism spectrum disorders. In O. Khatib, V. Kumar, and G. Pappas, editors, *Experimental Robotics*, volume 54 of *Springer Tracts in Advanced Robotics*, pages 201–210. Springer Berlin / Heidelberg.

[3] H. Kozima, C. Nakagawa, and Y. Yasuda. Interactive robots for communication-care: a case-study in autism therapy. In *Robot and Human Interactive Communication, 2005. ROMAN 2005. IEEE Int. Workshop on*, pages 341 – 346, 2005.

[4] K. Murphy, M. Schmidt, and K. Swersky. Conditional random field (CRF) toolbox for Matlab. http://www.cs.ubc.ca/~murphyk/Software/CRF/crf.html, 2010.

[5] P. Perego, S. Forti, A. Crippa, A. Valli, and G. Reni. Reach and throw movement analysis with support vector machines in early diagnosis of autism. In *Int. Conf. of the IEEE Engineering in Medicine and Biology Society, 2009.*, pages 2555 –2558, 2009.

[6] C. Rice. Prevalence of autism spectrum disorders. http://www.cdc.gov/mmwr/preview/mmwrhtml/ss5810a1.htm, 2009.

[7] B. Robins, K. Dautenhahn, and P. Dickerson. From isolation to communication: A case study evaluation of robot assisted play for children with autism with a minimally expressive humanoid robot. In *Second Int. Conf. on Advances in Computer-Human Interactions, 2009. ACHI '09.*, pages 205 –211, 2009.

[8] F. Shic, W. Jones, A. Klin, and B. Scassellati. Swimming in the underlying stream: Computational models of gaze in a comparative behavioral analysis of autism. In *in 28th Annual Conf. of the Cognitive Science Society*, pages 780–785, 2006.

[9] D. Xu, J. Gilkerson, J. Richards, U. Yapanel, and S. Gray. Child vocalization composition as discriminant information for automatic autism detection. In *Engineering in Medicine and Biology Society. Annual Int. Conf. of the IEEE*, pages 2518 –2522, 2009.