25th IEEE International Symposium on
Robot and Human Interactive Communication (RO-MAN)
August 26-31, 2016. Columbia University, NY, USA

Human Robot Activity Classification based on Accelerometer and Gyroscope*

Beibin Li[1], Laura Boccanfuso[1], Quan Wang[1], Erin Barney[1], Yeojin Amy Ahn[1], Claire Foster[1], Katarzyna Chawarska[1], Brian Scassellati[2], and Frederick Shic[1,2]

***Abstract*— Accelerometers have been widely used to record and classify human daily activities such as walking, sitting, and playing sports. However, these sensors have less often been used to classify Human-Robot Interaction (HRI), and rarely in the context of HRI with special populations. This paper uses triaxial accelerometers and gyroscopes embedded in a commercially available robot, Sphero, to classify four classes of human-robot interaction: kick, drop, hold, and no interaction. Fifteen children with Autism Spectrum Disorder (ASD) or Broader Autism Phenotype (BAP) played with Sphero freely for 10 minutes, and three typically developing (TD) adults interacted with Sphero under guided instructions to provide additional contextual ground truth. Accelerometer and gyroscope data were recorded, and video-taped sessions manually behaviorally coded. Thirty-six features were selected from sensor signals and tested using four supervised learning algorithms. The best model accuracy of this 4-class classification problem was obtained using random forest algorithm: for children the accuracy was 48.82%; for adults the accuracy reached 73.15%.**

I. Introduction

Sphero is a 3-inch diameter spherical robot with a high-impact polycarbonate plastic shell. It contains an internal guidance system that includes a gyroscope, accelerometer, and multi-colored LED lights. Sphero can roll around, vibrate, and change colors. Boccanfuso et al. have used Sphero in their prior studies [1][2] and showed that children with ASD with higher relative verbal ability avoided interacting with Sphero more when negative emotions were simulated.

While most research involving robots and children has been conducted in controlled laboratory settings where behavioral coding techniques can be used to manually annotate and categorize behaviors, these techniques are not scalable or easily accessible, creating a barrier to the use of robots for monitoring development in real-world daily lives. Similarly, using infrared light or computer vision techniques to classify HRIs at homes and schools are not secure nor affordable. Thus, we aimed to develop an autonomous system that could identify when and how participants were interacting with Sphero. Developing a classification algorithm to classify HRI interactions would not only provide alternative solution for video coding, but also provide real time feedback for use with online algorithms. With such a classification algorithm, scientists could further improve robot designs to assist in the phenotyping and identification of disease and atypical development, providing personalized social or cognitive robot-assisted interventions by dynamically adapting to users in real time. Salter et al. used similar sensors to classify human interactions [5], but only had 10% accuracy in detecting general interaction (e.g. kicking, pushing, and banging) from being alone, carried, and spun.

However, there are three challenges in HRI classification algorithm: 1. unlike human daily activity classification where only human movement is encountered, both humans and robots may move in HRI classification problem; 2. adults and children play differently—therefore we need different models for different populations; 3. from our observations, different children with ASD play differently with Sphero.

Figure 1. A child plays with Sphero in an examination room, while an experimenter and a parent accompany him.

II. Participants and Methods

A. Participants

Fifteen children (age M=31.31 months, SD=12.04 months, female=3, male=12) were included in the analyses, 13 diagnosed with ASD, and two identified with BAP. Each child was asked to play with Sphero freely in a room for 10 minutes. Each session was videotaped and parental consent was obtained. After the experiment, three coders manually coded drops, kicks, holds, and pick up actions in the videos.

Participants included three additional adults (*female=1, male=2*). Each adult participated in three, five-minute sessions, and in each session the participants were asked to drop, hold, or kick Sphero multiple times. Even though adults play differently than children (e.g. the height of a drop action), we evaluated the performance of interaction classifiers using adult training data to predict corresponding child interactions because of the relative ease of collecting data with adults and similarity of movement characteristics.

*Research supported by NSF Expedition in Socially Assistive Robotics 08-568, NIH awards K01 MH104739, R21 MH102572, R01 MH100182, R01 MH087554, DOD W81XWH-12-ARP-IDA, the FAR Fund, Hilibrand Family Foundation, Nancy Taylor Foundation for Chronic Diseases.

[1]Child Study Center/[2]Department of Computer Science, Yale University (e-mail: beibin.li at yale.edu, laura.boccanfuso at yale.edu, first.last at yale.edu, frederick.shic at yale.edu).

B. Data Analysis

Based on previous human activity classification study [3], a window size of 2 seconds was used for feature extraction. As illustrated in Fig. 2, high frequency and long magnitude changes were exhibited in the signals of kick and drop, and patterns of both differences and similarities were found between children and adults.

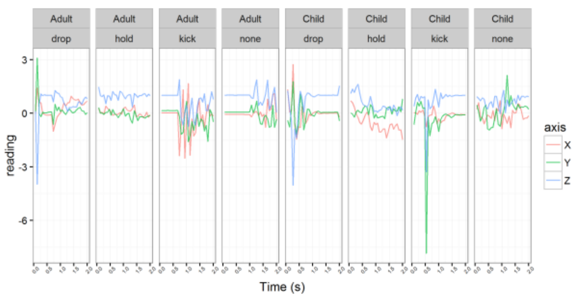

Figure 2. Sample Accelerometer Signals for Adults and Children

The mean, standard deviation, minimum, and maximum of signal readings within windows were used as time-domain features. We also took features from two domains in the study conducted by Bao et al., which used discrete fast Fourier transform (DFT) signal processing features [3]: sum of squares DFT component magnitudes as energy features and the normalized information entropy of the DFT component magnitudes as frequency-domain entropy. The four time-domain features, energy feature, and entropy feature were calculated for each axis of the accelerometer and gyroscope (6 channels total), and the resultant 36 features were used to train machine learning models.

Unstructured human robot activity (i.e. free play) is an emerging field where HRI taxonomy has not fully developed yet, and the physical proximity category is the closest for Sphero-human interactions. So, we classify four types of activity: holding, dropping, kicking, and no interaction, where the first three interactions are emphasized in Boccanfuso and her colleagues' work [3], and the last one is useful to determine whether a human is interacting with robot.

We examined the performance of four supervised learning algorithms: Support Vector Machines (SVM), Naïve Bayes Classifiers (NB), Random Forests (RF), and Classification Trees (CT). SVM with linear kernel and NB with kernel distribution were used in the analysis. We ran RF three times with either 2, 19, or 36 variables considered in each split, but the results of these three different split methods were similar. Four-class CT was also adapted for comparison.

Leave one subject out cross validation (LOSOCV) was used for all of the algorithms, such that data from one subject was left out as a test set and data from all other subjects were used as the training set. Bootstrapping was also used to improve stability for all four classification algorithms.

III. Results

As shown in Table I, the accuracy for the adult model was higher than that of children, and the accuracy of Random Forest algorithm was slightly higher than those of other algorithms. The Cohen's kappa value and unweighted average recall (UAR) also demonstrated the better performance of Random Forest compared to other algorithms.

Applying the adult model to predict children's interaction was less accurate than using children's model to predict children's interaction, suggesting that separate models for adults and children should be used.

TABLE I. Machine Learning Results

Algorithm	*Dataset*	*Accuracy*	*Kappa*	*UAR*
SVM	Child	32.44%	0.0951	0.2984
	Adult	59.85%	0.4154	0.5136
NB	Child	41.09%	0.1677	0.3800
	Adult	68.68%	0.5384	0.6139
RF	Child	48.82%	0.2790	0.4133
	Adult	73.15%	0.5936	0.6151
CT	Child	25.38%	0.0518	0.2476
	Adult	54.48%	0.3238	0.3528

IV. Conclusion

Different from previous works [1][5], this study used time-domain, frequency-domain, and energy features to classify broader and similar human robot interactions in an unstructured setting. Accelerometer and gyroscope signals from robots are sufficient to classify adult-robot interactions but inadequate for child-robot interaction classification. The classification accuracy was about 49% with data from fifteen children: significantly better than chance, but suggesting considerable room for improvement.

Random Forest performed slightly better than other three algorithms. It is important to note that our testing strategy was conservative: model accuracy could be improved by about 5% if we randomly chose 80% data for the training set without consideration of participant (c.f. the LOSOCV method). This finding highlights the need to consider multiple sources of variation in behavior identification training.

Many robot development tools allow developers to access motor data, which could significantly improve HRI classification models if researchers compare the differences between motor behavior and sensor data. We will stream motor signals, record more data, and improve learning algorithms in future work.

References

[1] L. Boccanfuso, E. S. Kim, J. C. Snider, Q. Wang, C. A. Wall, L. DiNicola, G. Greco, F. Shic, B. Scassellati, L. Flink et al., "Autonomously detecting interaction with an affective robot to explore connection to developmental ability," in *Affective Computing and Intelligent Interaction (ACII), 2015 International Conference on. IEEE*, 2015, pp. 1–7.

[2] L. Boccanfuso, E. Barney, C. Foster, Y. J. Ahn, K. Chawarska, B. Scassellati, F. Shic. "Emotional robot to explore differences in play patterns of children with and without ASD." *ACM/IEEE International Conference on Human-Robot Interaction*, 2016, pp. 19-26.

[3] L. Bao and S. S. Intille, "Activity recognition from user-annotated acceleration data," in *Pervasive computing*. Springer, 2004, pp. 1– 17.

[4] S. J. Preece, J. Y. Goulermas, L. P. Kenney, and D. Howard, "A comparison of feature extraction methods for the classification of dynamic activities from accelerometer data," *Biomedical Engineering, IEEE Transactions on*, vol. 56, no. 3, pp. 871–879, 2009.

[5] T. Salter, F. Michaud, D. Letourneau, D. C. Lee, I. P. Werry. "Using proprioceptive sensors for categorizing human-robot interactions," in *Human-Robot Interaction (HRI), 2007 2nd ACM/IEEE International Conference* on. IEEE, 2007, pp. 105-112.