Transfer Learning of Engagement Recognition within Robot-Assisted Therapy for Children with Autism

Nazerke Rakhymbayeva, Anara Sandygulova
Nazarbayev University
53, Kabanbay Batyr Ave., Nur-Sultan, Kazakhstan, 010 000
nazerke.rakhymbayeva@nu.edu.kz

Abstract

Social robots deployed in the therapy of autism is a promising and important research domain. Recently, an increasing amount of work is being conducted utilizing a social robot as a mediator between a therapist and a child with autism. Being able to evaluate how engaged a child is both offline and in real-time would improve the quality of the provided robot-assisted intervention and also provide objective metrics for later analysis by the therapist. The state-of-the-art engagement recognition is challenged by the diverse styles of expressing engagement by this vulnerable population group. To this end, this PhD project aims to explore how transfer learning can improve the recognition accuracy of children's engagement with the robot or another human. We will utilize four publicly available multi-modal datasets to discover a suitable feature representation of engagement during various types of activities with the robot.

Introduction

Social robots are increasingly being used as a mediator between a therapist and a child investigating their effect on the therapy of autism. Several projects have explored the utility of robots within autism therapy by measuring children's engagement with the robot during human-robot interaction studies (Kim et al. 2012). Such studies are typically video-recorded for later examination and annotation. Researchers would then manually video code the recordings using the ELAN software.

Engagement is defined as one's active and appropriate involvement in the task with the robot and/or the therapist. For example, Kim et al. (2012) and Rudovic et al. (2017) coded engagement scores on a 1-5 scale with 1 corresponding to being fully non-compliant (evasive) and 5 to be fully engaged. Kim et al. (2012) coded 10-sec-long fragments of videos while Rudovic et al. (2017) coded the whole engagement episode to preserve the context i.e. starting with the target task until one of the engagement scores is met. In our work, we coded engagement scores relative to the timing of the robot activities (e.g. dance, song, storytelling activities lasting for 1-3 minutes) (Sandygulova et al. 2019). While such manual video-coding produces reliable results, it requires expert knowledge and is extremely time-consuming.

In contrast, data-driven approaches learn from previous data examples and allow to speed up the process and also aim for real-time performance. In this project, we propose to take a data-driven approach to classify the engagement of autistic children with social robots or other humans utilizing at least four existing datasets.

To this end, this research work aims to first discover a suitable feature representation of engagement of an autistic child with a robot and another human in various types of activities with the robot by considering the multi-modality of the available datasets (audio, face, and body features). There are various research questions that we aim to address:

- Which multi-modal features are transferable from the typical population to the autistic population?
- Are there common features that discriminate children's engagement in human-human and human-robot scenarios?
- Since the nature of autism is heterogeneous (from mild to severe cases) and each child is unique, to what extend a child independence can be achieved?
- How generalizable a social interaction with another agent can be (i.e. applied to various types of robot activities)?

The Research Approach and Methodology

State of the art in Engagement Recognition

Human behavior is naturally multi-modal, and people use cues, such as eye gaze, hand gestures, expressions of emotions, body posture, and tone of voice to manage engagement and coordinate social interactions. Unfortunately, the performance of an engagement assessment is limited due to heterogeneous data. Past work presented results of engagement recognition on their own solicited datasets (Rudovic et al. 2017). However, their datasets are not publicly available due to ethical concerns of the vulnerable population group. Therefore, the first objective of this PhD project was to create a dataset by utilizing video recordings of the past robot-assisted autism therapy studies conducted with a total of 36 children. All video recordings were annotated by two independent raters who coded engagement scores from 0 to 5 relative to the timing of the applications and had an agreement score on 20% of cross-coded data computed from pair-wise ICC of the coders equal to 82.6%. The total number of hours of coded video was 48 hours and 34 minutes. We utilized the OpenPose library to extract the keypoints of the child from the video frames obtaining 2D information of 25 keypoints (joints) in a body and feet, 2x21 key points in both hands, and 70 keypoints in a face (Cao et al. 2019). We achieved 79.34% of accuracy on this dataset in a child independent split when taking into account body and face features for binary classification. However, for multi-class classification the model performance decreased almost by two times and equals 47.64%.

Publicly Available HRI Datasets

In order to improve the performance of engagement recognition, the second objective of this research is to explore existing publicly available HRI datasets, such as DREAM (Billing et al. 2020), PInSoRo (Lemaignan et al. 2018), and MHHRI (Celiktutan, Skordos, and Gunes 2017) in terms of the type of data provided (video, audio, depth, OpenPose keypoints) and available labels and annotations. The next step is to extract relevant features from each dataset and to train and test engagement recognition for each dataset. The next step is to explore transfer learning from one dataset to another.

The Transfer Learning Challenge

The next objective of this PhD project is to apply transfer learning to improve the performance of the state-of-the-art results in engagement recognition. We will evaluate various combinations of the primary dataset (source domain) and secondary dataset (target domain). The main challenge of using multiple datasets is that primary and secondary datasets are typically in vastly different formats. Datasets have their own set of instructions that need to be followed for the initial setup. Therefore, a lot of custom coding is necessary to transform and load data into a format that is readily usable by the model.

Optimal Feature Selection for Multi-modal Engagement Recognition

An optimal feature selection technique will be applied to work with multi-modal datasets. This approach will combine different kinds of data representations, such as audio, body, face, and other behavioral data of children in order to answer our RQs.

Contributions

During this PhD project the following contributions are expected:

- A dataset of behavioral data recorded from 36 children diagnosed with ASD. The dataset covers 194 therapy sessions and more than 48 hours of video (done).
- Experiments on the dataset to compare face and body modalities as well as two types of split: Child Independent and Session Independent (done).
- Applying transfer learning to four datasets by first acquiring knowledge from the primary dataset (source domain) to the secondary dataset (target domain) by assuming that each dataset has common features. The proposed architecture consists of input, some hidden, and output layers. The input layer consists of one of the dataset input features. The hidden layers use Relu and Softmax activation functions. The output layer consists of sigmoid function and cross entropy loss function for predicting the output class (from 0 to 5) (in progress).

Acknowledgments

I would like to thank my external supervisor, Prof. Wafa Johal from University of New South Wales in Australia for her invaluable advice and guidance for this research. This research is supported by the Nazarbayev University Collaborative Research Program grant (award number is 091019CRP2107).

References

Billing, E.; Belpaeme, T.; Cai, H.; Cao, H.-L.; Ciocan, A.; Costescu, C.; David, D.; Homewood, R.; Hernandez Garcia, D.; Gómez Esteban, P.; Liu, H.; Nair, V.; Matu, S.; Mazel, A.; Selescu, M.; Senft, E.; Thill, S.; Vanderborght, B.; Vernon, D.; and Ziemke, T. 2020. The DREAM Dataset: Supporting a data-driven study of autism spectrum disorder and robot enhanced therapy. *PLOS ONE* 15(8): 1–15. doi:10.1371/journal.pone.0236939. URL https://doi.org/10.1371/journal.pone.0236939.

Cao, Z.; Hidalgo Martinez, G.; Simon, T.; Wei, S.; and Sheikh, Y. A. 2019. OpenPose: Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence* .

Celiktutan, O.; Skordos, E.; and Gunes, H. 2017. Multimodal Human-Human-Robot Interactions (MHHRI) Dataset for Studying Personality and Engagement. *IEEE Transactions on Affective Computing* doi:10.1109/TAFFC.2017.2737019.

Kim, E.; Paul, R.; Shic, F.; and Scassellati, B. 2012. Bridging the Research Gap: Making HRI Useful to Individuals with Autism. *Journal of Human-Robot Interaction* 1. doi:10.5898/jhri.v1i1.25.

Lemaignan, S.; Edmunds, C. E. R.; Senft, E.; and Belpaeme, T. 2018. The PInSoRo dataset: Supporting the data-driven study of child-child and child-robot social dynamics. *PLOS ONE* 13(10): 1–19. doi:10.1371/journal.pone.0205999. URL https://doi.org/10.1371/journal.pone.0205999.

Rudovic, O. O.; Lee, J.; Mascarell-Maricic, L.; Schuller, B. W.; and Picard, R. W. 2017. Measuring Engagement in Robot-Assisted Autism Therapy: A Cross-Cultural Study. *Front. Robot. AI* 4(36). doi:10.3389/frobt.2017.00036. URL https://doi.org/10.3389/frobt.2017.00036.

Sandygulova, A.; Zhexenova, Z.; Tleubayev, B.; Nurakhmetova, A.; Zhumabekova, D.; Assylgali, I.; Rzagaliyev, Y.; and Zhakenova, A. 2019. Interaction design and methodology of robot-assisted therapy for children with severe ASD and ADHD. *Paladyn, J. Behav. Robot.* 10: 330–346.