Measuring Engagement Online in CRI

Pieter Wolfert
Tilburg University
Tilburg, The Netherlands
a.p.wolfert@uvt.nl

Paul Vogt
Tilburg University
Tilburg, The Netherlands
p.a.vogt@uvt.nl

Mirjam de Haas
Tilburg University
Tilburg, The Netherlands
mirjam.dehaas@uvt.nl

Pim Haselager
Radboud University
Nijmegen, the Netherlands
w.haselager@donders.ru.nl

The Near Future of Children's Robotics, IDC Workshop 2018, Trondheim, Norway

Abstract
In this paper, a pipeline is suggested for measuring child engagement in a robot tutoring task, together with a pilot experiment for verification. Smiling, gaze direction and posture are taken as indicators for engagement. A pilot experiment is proposed to test the performance of the model. This will be a robot tutoring task based on the child game "I spy with my little eye" during which children with the age of five learn English names for animals [6]. In this pilot experiment, the children are provided breaks when they are dis-engaged, to re-engage the children. Afterwards, the children will be asked to rate the perception of the robot, and it is expected that this rating will be higher than a robot without engagement detection.

Author Keywords
Child Robot Interaction; Machine Learning; Applied Robotics; Robot Tutoring

ACM Classification Keywords
· **Human-centered computing → HCI design and evaluation methods;** [· **Human-centered computing → Human computer interaction (HCI)**]

Introduction

In order to make tutor robots more effective, robots should be able to adapt to a user. In a study done by Leyzberg et al., [10] where students had to solve a grid-based logic puzzle, it was found that personalized tutoring improved the students' performance. Ramachandran et al. [12] found that providing personal breaks based on the number of errors during a task in a robot tutoring session improved learning in children. Improving the effectiveness of a robot tutoring task can be done in several ways. One way is to keep track of a child's progress over time, by registering the words that a child learnt [6, 12]. In a previous experiment of De Wit et al. [6] was found that adaptive tutoring did not have a significant effect on learning gain, but the use of gestures did have a significant effect. An important finding, however, was that engagement seemed to lower at the end of the interaction, in all conditions. With this in mind, another way is to improve the quality of the time a child spents with the robot. This can be done by adjusting the robots behavior such that it creates a better environment for interaction and tutoring, based on the child's engagement [14].

A robot that can detect a child's (lack of) engagement online during the interaction and respond to this appropriately could improve the quality of tutoring by re-engaging the child when so desired. It has been found that eye-gaze is a good indicator of engagement [8], and eye-gaze is proposed as a metric for engagement in human robot interaction [15, 1]. Body posture and head position are also identified as indicators for engagement [1], head pose is also used for realtime attention assessment [9]. Another predictor for engagement is smiling [5, 15]. Serholt et al. [15] found that smiles were most common after positive feedback. Measuring smiles can be done by using a multilayer convolutional neural network trained on facial expressions [2].

In this paper, a pipeline is proposed to measure engagement based on smiling, posture and gaze of a child during a robot tutoring task; such that an online prediction can be generated given a frontal video recording. In order to verify whether the pipeline works, a pilot experiment is suggested based on the experiment done by de Wit et al. [6].

This project is embedded in the L2TOR project. The L2TOR project ('el tutor') aims to design a social robot that supports the teaching of a second language to preschool children. This platform, which runs on a NAO humanoid robot, requires that the robot is able to work together with young children (age 5), on a peer level, and can provide relevant feedback.

Methods

The aim of this project is to construct and train a model to do online engagement prediction. This model will consist of three different neural networks (one for each feature), with a combined output resulting in a prediction. Furthermore, a method is developed to re-engage the child when the child seems disengaged. This method will then be tested and verified in a pilot experiment based on de Wit et al. [6], in a NAO humanoid robot. In the current L2TOR experiments the robot does not account for the child's engagement. By accounting for the engagement of a child with the robot, the behavior of the robot can be more matched to the child's behavioral state, by for example providing a break when the engagement seems to drop. Engagement will be measured by combining smiling, posture and gaze predictions. Two datasets will be used for engagement prediction: a dataset of a previous L2TOR study [6] and the EmoReact dataset [11]. The L2TOR dataset contains frontal video clips of subjects performing a language learning task with the robot. This dataset has been annotated with engagement observations, in the range low, medium, high. In each frame the face and the upper body of the child is visible.

The EmoReact dataset contains images annotated with emotional states of children between the ages 4 and 14. There are in total 17 states annotated. A subset will be used such that only 4 and 5 year old children are considered. The L2TOR dataset provides the engagement observations and the EmoReact dataset images with emotional expressions of children, useful for detecting smiles. The smiling prediction will be done using a convolutional neural network based on the model proposed by Arriaga et al. [2].

OpenPose [4] will be used to extract postures from the L2TOR dataset, this information is then, together with the engagement labels, used to train a model that can classify engagement. In order to do gaze prediction, the L2TOR data will be processed to identify both the position of the robot and the tablet (the tablet is part of the tutoring experiments). To perform gaze prediction a pretrained model provided by Recasens et al. [13] is used. This model provides the gaze direction of a child given the location of the face of the child in the video. Localisation of the face of the child will be done using OpenCV [3]. The three different neural networks' outputs are followed by an LSTM layer [7] –to make sure that temporal relations are learned– necessary for measuring engagement. Once this model is trained, it can be used online during an experiment and provide engagement predictions.

In the pilot experiment to evaluate the effectiveness of the model, participants (children) will learn six English animal words during a game of "I spy with my little eye". During the task a NAO humanoid robot will provide an animal name in English, after which the child has to select the correct animal on a tablet. In this pilot experiment, the robot will offer breaks during the experiment based on the engagement of the child, after this break the session continues. We expect that the engagement will increase after the break, based on the experiment of Ramachandran [12].

Conclusion

In this paper a model for online measuring of children's engagement in a robot tutoring task based on smiling, gaze and posture is proposed. A valid model for online measuring of engagement is not available yet, while such a model could influence learning in robot tutoring and can stimulate the developments of robot tutoring systems. The model and methods can be adjusted to work with other domains as well, outside the scope of robot tutoring, enhancing measuring of engagement in human robot interaction. The proposed model is currently being developed, and the expectation is that preliminary results of both the model and the evaluation pilot will be presented at the CRI-IDC Workshop.

Acknowledgments

This work has been supported by the EU H2020 L2TOR project (grant 688014).

REFERENCES

1. Salvatore M Anzalone, Sofiane Boucenna, Serena Ivaldi, and Mohamed Chetouani. 2015. Evaluating the engagement with social robots. *International Journal of Social Robotics* 7, 4 (2015), 465–478.

2. Octavio Arriaga, Matias Valdenegro-Toro, and Paul Plöger. 2017. Real-time Convolutional Neural Networks for Emotion and Gender Classification. *arXiv preprint arXiv:1710.07557* (2017).

3. Gary Bradski. 2000. The OpenCV Library. *Dr. Dobb's Journal of Software Tools* (2000).

4. Zhe Cao, Tomas Simon, Shih-En Wei, and Yaser Sheikh. 2017. Realtime multi-person 2d pose estimation using part affinity fields. In *CVPR*, Vol. 1. 7.

5. Ginevra Castellano, André Pereira, Iolanda Leite, Ana Paiva, and Peter W McOwan. 2009. Detecting user engagement with a robot companion using task and social interaction-based features. In *Proceedings of the 2009 international conference on Multimodal interfaces*. ACM, 119–126.

6. Jan de Wit, Thorsten Schodde, Bram Willemsen, Kirsten Bergmann, Mirjam de Haas, Stefan Kopp, Emiel Krahmer, and Paul Vogt. 2018. The Effect of a Robot's Gestures and Adaptive Tutoring on Children's Acquisition of Second Language Vocabularies. In *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*. ACM, 50–58.

7. Sepp Hochreiter and Jurgen Jürgen Schmidhuber. 1997. Long short-term memory. *Neural Computation* 9, 8 (1997), 1–32.

8. Ryo Ishii, Yukiko I Nakano, and Toyoaki Nishida. 2013. Gaze awareness in conversational agents: Estimating a user's conversational engagement from eye gaze. *ACM Transactions on Interactive Intelligent Systems (TiiS)* 3, 2 (2013), 11.

9. Séverin Lemaignan, Fernando Garcia, Alexis Jacq, and Pierre Dillenbourg. 2016. From real-time attention assessment to with-me-ness in human-robot interaction. In *The Eleventh ACM/IEEE International Conference on Human Robot Interaction*. IEEE Press, 157–164.

10. Daniel Leyzberg, Samuel Spaulding, and Brian Scassellati. 2014. Personalizing robot tutors to individuals' learning differences. *Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction - HRI '14* March 2014 (2014), 423–430.

11. Behnaz Nojavanasghari, Tadas Baltrušaitis, Charles E Hughes, and Louis-Philippe Morency. 2016. EmoReact: a multimodal approach and dataset for recognizing emotional responses in children. In *Proceedings of the 18th ACM International Conference on Multimodal Interaction*. ACM, 137–144.

12. Aditi Ramachandran, C.-M. Huang, and Brian Scassellati. 2017. Give Me a Break!: Personalized Timing Strategies to Promote Learning in Robot-Child Tutoring. *ACM/IEEE International Conference on Human-Robot Interaction* Part F1271 (2017), 146–155.

13. Adria Recasens, Aditya Khosla, Carl Vondrick, and Antonio Torralba. 2015. Where are they looking?. In *Advances in Neural Information Processing Systems*. 199–207.

14. Thorsten Schodde, Kirsten Bergmann, and Stefan Kopp. 2017. Adaptive Robot Language Tutoring Based on Bayesian Knowledge Tracing and Predictive Decision-Making. *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction - HRI '17* (2017), 128–136.

15. Sofia Serholt and Wolmet Barendregt. 2016. Robots tutoring children: Longitudinal evaluation of social engagement in child-robot interaction. *ACM International Conference Proceeding Series* 23-27-Octo (2016).