Mixing Implicit and Explicit Probes: Finding a Ground Truth for Engagement in Social Human-Robot Interactions

Lee J. Corrigan
University Of Birmingham
United Kingdom
ljc228@bham.ac.uk

Christina Basedow
Jacobs University Bremen
Germany
c.basedow@jacobs-university.de

Dennis Küster
Jacobs University Bremen
Germany
d.kuester@jacobs-university.de

Arvid Kappas
Jacobs University Bremen
Germany
a.kappas@jacobs-university.de

Christopher Peters
Royal Institute of Technology (KTH), Sweden
chpeters@kth.se

Ginevra Castellano
University Of Birmingham
United Kingdom
g.castellano@bham.ac.uk

ABSTRACT

In our work we explore the development of a computational model capable of automatically detecting engagement in social human-robot interactions from real-time sensory and contextual input. However, to train the model we need to establish ground truths of engagement from a large corpus of data collected from a study involving task and social-task engagement. Here, we intend to advance the current state-of-the-art by reducing the need for unreliable post-experiment questionnaires and costly time-consuming annotation with the novel introduction of implicit probes. A non-intrusive, pervasive and embedded method of collecting informative data at different stages of an interaction.

Categories and Subject Descriptors:
H.1.2 [Information Systems]: Human factors

Keywords: Ground Truth; Engagement; Social Tasks; Machine Learning; Human-Robot Interaction.

1. INTRODUCTION

The phenomenon of engagement is a widely discussed topic in the field of human-robot interaction (HRI), but in our work we wish to advance the current state-of-the-art by developing an on-line engagement detector. The detector should be capable of differentiating between engagement with the task and social engagement with the robot. Prior to our work, affect ground truths necessary to train automatic detectors have typically been determined post-experimentally using domain-specialist annotation. Here, we are exploring ways to find those truths from the users during the interaction in a manner that is non-intrusive. We have several explicit probes embedded into the interaction in terms of questionnaires, and complement these with implicit probes as task-related indicators allowing a real-time account of engagement states at the beginning, during and end of the interaction. Rather than including self-report only retrospectively at the end, we use a mixed-design with embedded questions and assessment throughout different stages of the task.

HRI'14, March 3–6, 2014, Bielefeld, Germany.
ACM 978-1-4503-2658-2/14/03.
http://dx.doi.org/10.1145/2559636.2559815.

2. RELATED WORK

To train our computational model, we first need to establish ground truths for the types of engagement that we aim to measure empirically. Typically, the self-report data needed to train this type of computational model is expensive and time-consuming to collect, and often requires trained annotators who can reliably classify behavioural data according to the construct that is to be measured. Whilst many attempts have been made to record affect ground truths, no unambiguous method currently exists [1]. Typical approaches involve post-experiment annotation of continuous or segmented videos of the recorded interaction between the user and the system [2].

In HRI research, the concept of engagement is frequently seen as a binary concept, relating to whether or not someone is engaged in the interaction as a whole. Often there is no distinction between the different aspects of the interaction or how one aspect influences engagement with another. However, in this project, engagement is seen as a far more developed phenomenon which builds upon work in social engagement, relating to the connection between the human and the robot during an interaction [3], and task engagement, considered to be characterised by elements of attention, concentration and enjoyment [4]. Further still, we view engagement in a social task as the fusion of both social and task engagement.

3. SCENARIO

Our experimental scenario involves three interactive game-like tasks [5]. The first two tasks are designed to elicit maximally different states of task engagement and the final task is designed to elicit social-task engagement. The technical implementation of our scenario is comprised of a large touch-screen table to graphically represent the interactive tasks, a torso-only version of the *NAO* humanoid robot to facilitate the social aspect of the interaction, several video cameras

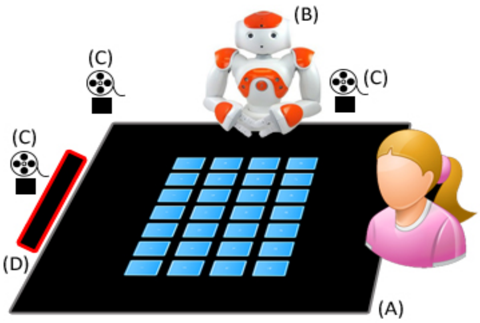

Figure 1: Overview of the technical set up: (A) Touch Screen Table, (B) NAO Robot (C) Cameras (D) Microsoft Kinect. Human character courtesy of http://www.icons-land.com.

for detecting facial expressions, a *Microsoft Kinect* for gaze direction and lean position relating to posture, and an *Affectiva Q Sensor* for measuring galvanic skin response.

4. SOCIAL TASK

For clarification, in a HRI context a social task can be described as a task performed in collaboration with a social robot. The task is an explicit task where the output is directly related to the inputs provided. Here, the robot's behaviour, timing and utterances can influence the partner's effectiveness within the task, likewise, the flow of the task can influence the perceived value of collaborating with the robot.

5. GROUND TRUTH

Using our experimental scenario, we captured a corpus of real-time multi-modal sensor readings and task-related context from eighty participants who took part in a recent HRI study. By using the actual responses to the implicit and explicit probes used in that study, we intend to support our claims by finding the least ambiguous datasets for training our model of engagement. For the purpose of critical evaluation, both post-experiment statistical analysis and traditional domain-specialist annotation will be used to validate our probes and to help find ground truths for engagement.

6. EXPERIMENTAL PROBES

A probe can be a non-intrusive, pervasive and embedded method of collecting informative data at different stages of an interaction. Here, the explicit probes (i.e., questionnaires) are embedded into the interaction itself, minimising disruption caused by researcher intervention. The implicit probes are pervasive, executing at predefined stages within the interaction, minimising unnecessary disruption to the natural interactive flow. The feedback we gain from these probes is used as ground truths for training our models as well as milestones for other methods of data mining and statistical analysis.

6.1 Implicit Probes

6.1.1 Social Bonding Probe

This implicit probe occurs at the start of the third task, allowing the participant to choose whether or not they wish to interact with the robot.

6.1.2 Robot to Task

The second probe is also an implicit probe designed to measure the temporal lag involved with diverting gaze from the robot to objects which are graphically represented within the task. The results of this probe tells us how much attention the participant gives to the robot's implied instruction.

6.1.3 Task to Robot

During the interaction the robot will ask the participant a question which requires the participant to divert gaze and attention away from the task, towards the robot. Here, we detect any shift of gaze towards the robot and measure the initial temporal lag and further sustained gaze.

6.1.4 Attention to Instructions

If at any point during the task, the participant presses on a square whilst another is open, a buzzer will sound and the robot will inform the participant that they risk damaging the system if they press a square before the previous one has recovered. Following a warning, we measure the temporal difference of any future warnings.

6.2 Explicit Probes

6.2.1 Regular Self-Report Probes

The repeated self-report is embedded within the interaction to minimise disruption. This on-screen self-report probe occurs every minute, allowing us to measure changes of task and social engagement at different stages of the interaction.

7. CONCLUSION

Mixing implicit and explicit probes allows us to derive ground truths from within the interaction itself, reducing the need for expensive domain-specialist annotation. These probes can be used to train computational models capable of classifying engagement from multi-modal datasets, offering a more reliable temporal view of the social and task state compared to that of traditional post experiment questionnaires.

Acknowledgement

This work was partially supported by the European Commission (EC) and was funded by the EU FP7 ICT-317923 project EMOTE (EMbOdied-perceptive Tutors for Empathy-based learning). The authors are solely responsible for the content of this publication. It does not represent the opinion of the EC, and the EC is not responsible for any use that might be made of data appearing therein.

8. REFERENCES

[1] J. Healey, "Recording affect in the field: Towards methods and metrics for improving ground truth labels," in *Proceedings of the 4th International Conference on Affective Computing and Intelligent Interaction - Volume Part I*, ser. ACII'11. Berlin, Heidelberg: Springer-Verlag, 2011, pp. 107–116.

[2] H. Gunes and B. Schuller, "Categorical and dimensional affect analysis in continuous input: Current trends and future directions," *Image and Vision Computing*, 2012.

[3] C. Rich, B. Ponsleur, A. Holroyd, and C. L. Sidner, "Recognizing engagement in human-robot interaction," in *5th ACM/IEEE International Conference on Human-Robot Interaction, HRI 2010*, 2010, pp. 375–382.

[4] D. J. Shernoff, M. Csikszentmihalyi, B. Schneider, and E. S. Shernoff, "Student engagement in high school classrooms from the perspective of flow theory," *School Psychology Quarterly*, vol. 18, no. 2, pp. 158–176, 2003.

[5] L. J. Corrigan, C. Peters, and G. Castellano, "Identifying task engagement: Towards personalised interactions with educational robots," in *Affective Computing and Intelligent Interaction (ACII), 2013 Humaine Association Conference on*, 2013, pp. 655–658.