https://doi.org/10.31256/Mm8Xb7O

Automatic Engagement Detection for Social Situation Assessment

Nicola Webb
Bristol Robotics Laboratory
University of the West of England
Bristol, United Kingdom
nicola.webb@brl.ac.uk

Manuel Giuliani
Bristol Robotics Laboratory
University of the West of England
Bristol, United Kingdom
manuel.giuliani@brl.ac.uk

Séverin Lemaignan
PAL Robotics
Barcelona, Spain
severin.lemaignan@pal-robotics.com

***Abstract*—In previous work, we collected simple interaction data through an online game. This collection method meant there was limited signal collection. From this, we created a metric to measure visual social engagement. In this work, we wish to collect an in-person dataset where we can capture a greater number of social signals. With this, we look to enrich our metric to investigate whether we can create more accurate profiles of engagement. Also this dataset looks to capture dynamic group interactions, both human-human and human-robot, with participants being asked to complete a series of tasks both individually and alone. From this data, we will create a model of social engagement to automatically detect engagement levels, in order to improve a robots social awareness.**

***Index Terms*—social interaction, social engagement, human interaction, social behaviour, human-robot interaction**

I. Introduction

When we enter a social situation, we can quickly interpret the signals around us and act accordingly. After the initiation of an interaction, there is the possibility of more interactions to follow, depending on the agents in the interaction and our interaction behaviours [1]. If a robot is looking at a group of humans interacting, a consistent data stream of all available signals for analysis may not be available, therefore leaving it difficult to assess the situation. For example, if one group member is occluded, understanding the group's overall interaction state will become challenging.

In this work, we look to collect a corpus of dynamic group interactions with both human-human and human-robot groups and also record several social signals present to model social engagement.

II. Social Situation Assessment

Situational awareness is the capacity to recognise the social signals in one's environment. Endsley [3] defined three levels of situational awareness: (1) perception, (2) interpretation and (3) evaluation. Humans are able to take several social signals (1), understand (2) and act accordingly/anticipate the direction of the current and potential interactions (3).

Fong et al. [6] surveyed the current social robot literature and stated that human-facing robots should be able to detect gestures, social cues and responses. In order for this to be possible, such robots will need to possess a level of social awareness; social robots at present can record various social signals in their environment. However, by simply extracting the social signals in an interactive environment, only the first level of situational awareness is reached. [7] explained that an artificial agents social cognition is limited by its ability to interpret its environment. For a social robot to attain the second and third levels of social awareness and reach reasonable level of social cognition, it will need to be able to process these signals.

III. Previous Work: Visual Social Engagement Metric

In previous work [4], we completed an online data collection of an online game. Participants played an online socially interactive game, simulating simple group interactions. From this, we created a visual social engagement metric which uses proxemics and mutual gaze. Figure 1 shows the visual social engagement metric for one participant interacting with one in-game NPC (non-player character).

From the metric, we created unique interaction profiles that show how participants interact, for example, how engaged is someone when beginning or ending an interaction. The profiles were created using the data 15 seconds preceding the initiation of interaction and the 15 seconds following. Once initiated, at frame 0, there is a spike in the level of engagement. These profiles are a way to measure how people start an interaction and can be used to learn patterns of interaction behaviours.

IV. Data Collection

Typically, social interactions are dynamic; interaction groups are formed and terminated based on context and interpersonal relations [5]. Therefore in this corpora, we wish to capture dynamic interactions with dynamic group formations of various sizes. To do so, participants will have to complete 'tasks' that will inadvertently influence them to talk to different participants, form groups or split up. The interpretation of the tasks will be up to the participant, so they are free to complete them as they see fit. The tasks are delivered through a web app accessed on their mobile devices. The groups will consist of 2 to 5 people, creating a variety of formations. Examples of the types of tasks are as follows:

– *Form a group of n players*
– *Bring x player to y location*

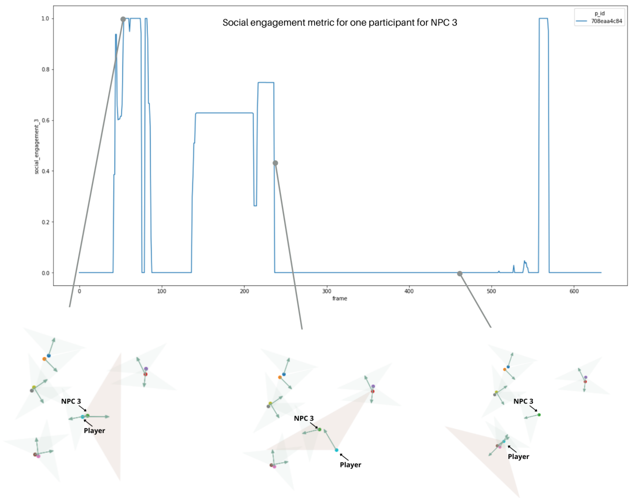

Fig. 1. Example visual social engagement metric for one participant interacting with an in-game NPC (NPC 3)

– *Bring x item to y location*

Some tasks will involve participants needing to interact with a Pepper robot (SoftBank Robotics). The tasks will see participants have to verbally ask questions to the robot controlled by a human operator (Wizard-of-Oz operation), so it would mirror the form of communication you would expect in the wild.

The tasks have a score value, which upon completion will be added to the participant's overall score. If they choose to skip a task they will be deduced points, to incentivise task completion. Participant scores will be recorded but will not form part of the engagement metric. Sessions will last approximately 15 minutes to collect enough data without reaching participant fatigue.

Before beginning the session, participants will provide basic demographic information and complete the Short IPIP-BFM-20 Questionnaire [2]. In the previous work mentioned, we did not find a correlation between the engagement profiles and personality, however, with the additional signals, we may see a positive correlation. If a correlation is found, this could suggest that personality types can predict our level of engagement in interactions.

During the game, we will be recording the participant's proximity information and gaze direction using a VICON motion capture system. Concurrently, facial and audio recordings taken using GoPros, will be taken of each participant. The audio recordings will not undergo any language processing but will be used to recognise whether a participant is speaking or not. Previously, only proximity and gaze direction data were collected, therefore the addition of facial and audio data will potentially improve the social engagement metric as we will have more opportunities mark a participant as presenting engaged social signals.

A. *Model of Engagement*

After data collection, we will develop a model that will detect the levels of engagement, using the signals collected as input. The model will be used to detect both the engagement levels of individual group members, as well as a group as a whole. By knowing how engaged those in an interaction are, it becomes easier to determine what the most appropriate next steps are in regards to one's own ways of interacting.

Additionally, as data is being collected in a temporal manner, it is possible to predict how someone's social behaviour will change over time. For example, once all group members' level of engagement starts decreasing, it is fair to assume that the interaction will soon come to an end and the group formation may break.

V. Conclusion

The proposed dataset will contain recordings of those social signals necessary for modelling social behaviours such as group dynamics, engagement and social attitudes towards robots. We believe this corpora to be of important to the wider research community as it will firstly capture dynamic human-human and human-robot interactions and secondly it will help to build our visual social engagement metric. We believe this metric to be a good contribution to improving the social awareness of a robot as knowledge of human engagement levels will help to build the robots understanding of its environment and guide the robots behaviour.

Potential uses of the corpora include building models for detecting social engagement or interpreting individual social behaviours and group dynamics. By giving social agents the ability to understand their environment and improve their social intelligence by increasing their social situation awareness, they will better be able to integrate into social scenarios and alter their behaviour accordingly.

References

[1] Smith, Mark. "Social situation, social behavior, social group." Psychological Review 52.4 (1945): 224.
[2] Topolewska, Ewa, et al. "The short IPIP-BFM-20 questionnaire for measuring the Big Five." Annals of Psychology 17.2 (2017): 385-402.
[3] Endsley, Mica R. "Toward a theory of situation awareness in dynamic systems." Situational awareness. Routledge, 2017. 9-42.
[4] Webb, N, et al. [No date]. Measuring Visual Social Engagement from Proxemics and Gaze. Unpublished.
[5] Keyton, Joann. "Group termination: Completing the study of group development." Small Group Research 24.1 (1993): 84-100.
[6] Fong, Terrence, Illah Nourbakhsh, and Kerstin Dautenhahn. "A survey of socially interactive robots." Robotics and autonomous systems 42.3-4 (2003): 143-166.
[7] Szczepanowski, Remigiusz, et al. "Computational Models of Consciousness-Emotion Interactions in Social Robotics: Conceptual Framework." Cognitive and Computational Neuroscience: Principles, Algorithms and Applications (2018): 79.