Continuous Interaction Data Acquisition and Evaluation

A Process Applied within a Smart, Robot Inhabited Apartment

Viktor Richter
Applied Informatics (CITEC)
Bielefeld University
Bielefeld, Germany
vrichter@techfak.uni-bielefeld.de

Franz Kummert
Applied Informatics (CITEC)
Bielefeld University
Bielefeld, Germany
franz@techfak.uni-bielefeld.de

ABSTRACT

Intelligent agents need to perceive and correctly interpret the social signals of their interaction partners. In order to support the development of these skills, we establish a process of long-term data acquisition, annotation and continuous model evaluation. We facilitate automatic recording and annotation of unconstrained, multicentric interactions in a smart environment. Finally, we simplify manual ground truth annotation and allow continuous evaluation of our recognition models on a growing set of interactions.

CCS CONCEPTS

• **Computing methodologies** → **Intelligent agents**; **Model verification and validation**; *Cognitive science*;

ACM Reference Format:
Viktor Richter and Franz Kummert. 2018. Continuous Interaction Data Acquisition and Evaluation: A Process Applied within a Smart, Robot Inhabited Apartment. In *HRI '18 Companion: 2018 ACM/IEEE International Conference on Human-Robot Interaction Companion, March 5–8, 2018, Chicago, IL, USA.* ACM, New York, NY, USA, Article 4, 2 pages. https://doi.org/10.1145/3173386.3177005

1 INTRODUCTION

Especially in long-term and multicentric interactions, socially intelligent agents need to understand their naïve interaction partners' behaviour and act according to their expectations. These expectations are mostly derived from own experiences in Human-Human Interaction (HHI). As a result, social robots can mimic human communication signals to e.g. claim the conversational floor [8] or affect the perceived conversational roles in an interaction [5]. Meanwhile, the models used to recognise human communication signals are often tailored to a specific scenario and seldom evaluated over a longer period.

The aim of this paper is to establish a process (see Figure 1 for an overview) and create the tools to automatically: (I) Provide a constantly growing set of unconstrained HHI and Human-Robot Interaction (HRI) interaction data, (II) automatically pre-annotate interactions using established models, and (III) continuously and incrementally quantify the performance of different models. Consequently, this simplifies the manual ground truth annotation and the design and evaluation of new models. Initially, we address models for conversational group detection and conversational role recognition.

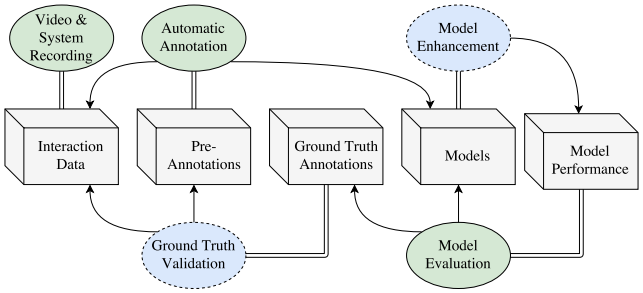

Figure 1: The data collection and evaluation process. Blocks depict data. Green circles depict automated parts. Blue circles (with dashed lines) highlight manual parts. Arrows depict *uses*-relations. Double lines show data production.

In HHI, focused encounters are often investigated in situations like meetings, seminars or spontaneous, free-standing interactions. Conversational groups are then detected using the speech activity of participants [1] or by applying the framework of F-Formations [4] and using body orientations [7]. In HRI it is often assumed that the group consists of the robot and the people it observes. Conversational roles are then recognised using heuristics – e.g. by detecting the current speaker and deriving the addressee from its gaze [6, 8]. Research on conversational groups often provides data sets (see [7]) but usually does not consider robotic encounters. Conversational role recognition in HRI usually leaves the problem of group detection aside and focuses on the generation and impact of the robot's behaviour in short, specific interactions. Finally, none of the presented works focuses on unconstrained, multicentric HRI.

To allow robots to interact in multicentric long-term scenarios, we need flexible and reliable models for conversational group detection and role recognition. By applying the proposed process, we create the environment necessary to create and enhance these models.

2 INCREMENTAL PROCESS

In this section we describe our data acquisition and model evaluation process (see Figure 1). We discuss its state of implementation and the possible benefits and biases of the chosen data acquisition strategy.

HRI '18 Companion, March 5–8, 2018, Chicago, IL, USA

ACM ISBN 978-1-4503-5615-2/18/03.
https://doi.org/10.1145/3173386.3177005

Figure 2: Excerpt from a first recording: *Upper-right*: top-down view of the kitchen. *Lower-right*: Flobi's perspective on the kitchen. *Rest*: overview recordings of the apartment

2.1 Data Collection & Annotation

For data acquisition, we use the Cognitive Service Robotics Apartment (CSRA)[1] smart home laboratory. It is designed to appear like a normal apartment, inhabited by multiple Flobi simulations and the robot Floka [9]. Furthermore, it is always operational and hosts activities like meetings, socialising events and demonstrations. This setup allows very natural and unrestricted HHI and HRI observations. During demonstrations, there are user groups from various backgrounds and age-groups. Participants of meetings and social events usually have an academical or technical background and some experience in HRI. The *video & system recording* phase (Figure 1) currently produces *interaction data* by recording: (I) system events and (II) overview videos (as in [3]), (III) videos from the Flobis' perspectives and finally (IV) twelve local top-down views covering the apartment. An excerpt from the video recordings can be seen in Figure 2. The recording automatically starts and stops when people enter and leave the apartment. As an initial test we recorded a prolonged demonstration of about 1 hour duration with eight Students, two lecturers and a presenter, where HHI and single- and multi-party HRI can be observed (see Figure 2).

In the *automatic annotation* phase, the recorded system data can be used to generate annotations of the apartment's and agent's beliefs (e.g. person positions, interaction times and dialog content). Additionally, we can analyse and annotate the recordings with computationally more complex tools which can not be used during the interaction. These *pre-annotations* can be imported to ELAN [2] and notably reduce the amount of work needed for manual *ground truth validation* (Figure 1). Because ELAN is not equally suitable for all annotation tasks (e.g. person positions and conversational groups), we are currently developing a specialised tool for this task.

2.2 Continuous Model Evaluation

We started with initial baseline *models* for group detection and conversational role recognition. The conversational group detection uses the F-Formations detection by [7] with some extensions[2]. The conversational role recognition uses Naive Bayes to classify the speaker, addressee and side-participants from the agent's point of view (similar to [6]). It uses the visual focus of attention and mouth movements of all participants observed by the agents or robot.

When an interaction is sufficiently annotated for the *model evaluation* phase, we add it to the evaluation data set. By versioning changes to the data and model, we can reiterate the evaluation process after each change and create a continuous assessment of the model in the *model performance* history (Figure 1). This facilitates the understanding of the impact to model changes and finally the *model enhancement* process.

[1] https://www.cit-ec.de/en/csra
[2] https://github.com/vrichter/fformation-rsb

3 CONCLUSION

We identified that understanding human social signals in HRI is a requirement to allow socially adequate behaviour and smooth interaction of robots and virtual agents. To achieve this, we proposed a process of incremental data acquisition and continuous model (re-)evaluation. The data acquisition was integrated into a smart, robot inhabited apartment, which exhibits diverse situations with mixed and multicentral HHI and HRI. We facilitated the annotation process and proposed a way to make the impact of new data or models changes directly visible through continuous evaluation.

By applying this framework in the proposed environment, we pave the way for high quality recognition models that can cope with various types of unrestricted, dynamic interactions. The choice of the models will allow a better understanding of human negotiation of conversational groups and roles and the involved processes.

4 ACKNOWLEDGMENTS

This work was supported by the Cluster of Excellence Cognitive Interaction Technology "CITEC" (EXC 277) at Bielefeld University, funded by the German Research Foundation (DFG).

REFERENCES

[1] Oliver Brdiczka, Jérôme Maisonnasse, and Patrick Reignier. 2005. Automatic detection of interaction groups. In *ACM International Conference on Multimodal Interaction (ICMI '05)*. https://doi.org/10.1145/1088463.1088473
[2] Hennie Brugman and Albert Russel. 2009. Annotating multi-media / multi-modal resources with ELAN. In *Int. Conf. on Lang. Resources and Evaluation (LREC '09)*.
[3] Patrick Holthaus, Christian Leichsenring, Jasmin Bernotat, Viktor Richter, Marian Pohling, Birte Carlmeyer, Norman Köster, Sebastian Meyer zu Borgsen, Birte Zorn, Rene Schiffhauer, Kai Frederic Engelmann, Florian Lier, Simon Schulz, Philipp Cimiano, Friederike Eyssel, Franz Kummert, Thomas Herrmann, David Schlangen, Ulrich Rückert, Sven Wachsmuth, Britta Wrede, and Sebastian Wrede. 2016. How to Address Smart Homes with a Social Robot? A Multi-modal Corpus of User Interactions with an Intelligent Environment. In *International Conference on Language Resources and Evaluation (LREC '16)*.
[4] Adam Kendon. 2010. Spacing and Orientation in Co-present Interaction. 1–15. https://doi.org/10.1007/978-3-642-12397-9_1
[5] Bilge Mutlu, Toshiyuki Shiwa, Takayuki Kanda, Hiroshi Ishiguro, and Norihiro Hagita. 2009. Footing in human-robot conversations. In *International Conference on Human Robot Interaction (HRI '09)*. https://doi.org/10.1145/1514095.1514109
[6] Viktor Richter, Birte Carlmeyer, Florian Lier, Sebastian Meyer zu Borgsen, Franz Kummert, Sven Wachsmuth, and Britta Wrede. 2016. Are you Talking to me? Improving the Robustness of Dialogue Systems in a Multi-party HRI Scenario by Incorporating Gaze Direction and Lip Movement of Attendees. In *Int. Conference on Human Agent Interaction (HAI '16)*. https://doi.org/10.1145/2974804.2974823
[7] Francesco Setti, Chris Russell, Chiara Bassetti, and Marco Cristani. 2015. F-Formation Detection: Individuating Free-Standing Conversational Groups in Images. *PLOS ONE* 10, 5 (2015). https://doi.org/10.1371/journal.pone.0123783 arXiv:arXiv:1409.2702v1
[8] Gabriel Skantze, Martin Johansson, and Jonas Beskow. 2014. Exploring Turn-taking Cues in Multi-party Human-robot Discussions about Objects. In *ACM International Conference on Multimodal Interaction (ICMI '14)*.
[9] Sebastian Wrede, Christian Leichsenring, Patrick Holthaus, Thomas Hermann, and Sven Wachsmuth. 2017. The Cognitive Service Robotics Apartment. *KI - Künstliche Intelligenz* 31, 3 (2017). https://doi.org/10.1007/s13218-017-0492-x