Understanding Social Behavior in Dyadic and Small Group Interactions

Preface

Cristina Palmero CRPALMEC7@ALUMNES.UB.EDU
Universitat de Barcelona and Computer Vision Center, Spain

Julio C. S. Jacques Junior JJACQUES@CVC.UAB.CAT
Computer Vision Center, Spain

Albert Clapés ALCL@CREATE.AAU.DK
Aalborg University, Denmark, and Computer Vision Center, Spain

Isabelle Guyon GUYON@CHALEARN.ORG
LISN (CNRS/INRIA) Université Paris-Saclay, France, and ChaLearn, USA

Wei-Wei Tu TUWEIWEI@4PARADIGM.COM
4Paradigm, China

Thomas B. Moeslund TBM@CREATE.AAU.DK
Aalborg University, Denmark

Sergio Escalera SERGIO@MAIA.UB.ES
Universitat de Barcelona and Computer Vision Center, Spain

Editors: Cristina Palmero, Julio C. S. Jacques Junior, Albert Clapés, Isabelle Guyon, Wei-Wei Tu, Thomas B. Moeslund, Sergio Escalera

Human interaction has been a central topic in psychology and social sciences, aiming at explaining the complex underlying mechanisms of communication from cognitive, affective and behavioral perspectives. From a computational point of view, research in dyadic and small group interactions is enabling the development of automatic approaches for detection, understanding, modeling, and synthesis of individual and interpersonal social signals and dynamics. Many human-centered applications for good (e.g., early diagnosis and intervention, augmented telepresence, and personalized assistive agents) depend on devising solutions for such tasks.

This special volume aims to bring together researchers in the field and from related disciplines to discuss advances and new challenges on understanding social behavior in dyadic and small group interactions, identify the main strengths and limitations of existing approaches, and define future research directions. In this context, we solicited papers in the form of tutorials, surveys, or novel technical and scientific contributions addressing the issues related, but not limited to: detection, understanding, modeling, prediction, and synthesis of individual and interpersonal social signals and dynamics; verbal and nonverbal communication analysis; contextual analysis; datasets; annotation frameworks and protocols; bias discovering and mitigation methods; interpretability and explainability.

Participants of the ChaLearn LAP Challenge on Understanding Social Behavior in Dyadic and Small Group Interactions (DYAD) were also invited to contribute to this volume

to present their proposed methodologies. The DYAD challenge and associated workshop were held in conjunction with the International Conference on Computer Vision (ICCV), 2021. Other researchers interested in contributing to this volume were invited to the workshop to briefly present their work as *sneak peek* presentations.

Invitations to contribute did not imply guaranteed paper acceptance. All submitted papers followed a rigorous peer-review process, and accepted papers were asked to carefully address the reviewers' comments to improve the final version of the papers. The selected 11 papers propose novel advances or discuss current approaches and potential future directions in the topics of personality recognition and perception, non-verbal behavior forecasting, body and head pose estimation, non-verbal synchrony, eye tracking, sentiment perception, and annotation frameworks. Each paper is described in short below.

The first paper, entitled "*Chalearn LAP challenges on self-reported personality recognition and non-verbal behavior forecasting during social dyadic interactions: dataset, design, and results*" summarizes the ChaLearn LAP DYAD'21 Challenge, the proposed baselines, and the results obtained by the top-performing solutions for each competition track: 1) self-reported personality recognition, and 2) behavior forecasting. The paper discusses the outcomes of each track, in addition to current related work and potential future research questions for each track topic. It also describes in detail the UDIVA v0.5 dataset, which was used for the challenge. The following five papers are associated to the challenge tracks. "*Learning personalised models for automatic self-reported personality recognition*", the winning approach of track 1, proposes to learn personalized models of self-reported Big-five personality traits using Neural Architecture Search (NAS), focusing on gender and age attributes. Results suggest that personalized models can improve the performance if compared to a generic model. In a post-challenge study, "*Skeleton-based personality recognition using Laban movement analysis*" presents a Graph Convolutional Network for estimating individuals' Big Five personality traits using a feature transformation based on Laban Movement Analysis, achieving higher performance than the state of the art without exposing image-level information to the network, which otherwise can leave the system susceptible to bias and result in ethical issues. Regarding track 2, "*Context-aware human behaviour forecasting in dyadic interactions*" introduces a generative framework with context awareness that captures the influence of the interacting partner's non-verbal signals on the target individual for behavior forecasting. Results show that the approach can efficiently capture the influence of the interacting partner's social signals on the target individual. The paper "*Comparison of spatio-temporal models for human motion and pose forecasting in face-to-face interaction scenarios*" explores the unimodal, dyadic, and multimodal capacities of several spatio-temporal models for human motion forecasting on UDIVA v0.5. Their Transformer-based architecture achieves state-of-the-art results on the novel dataset. In addition, the survey entitled "*Didn't see that coming: a survey on non-verbal social human behavior forecasting*" proposes a taxonomy that unifies the works from the social signal forecasting and the human motion forecasting fields, which have traditionally omitted each other. Given their many conceptual and methodological similarities, the authors argue that they may benefit from a joint consideration.

The remaining five papers consider aspects not associated to the challenge. The work entitled "*Head and body orientation estimation with sparse weak labels in free standing conversational settings*" focuses on estimating human head and body orientations in free-

standing conversational settings. The study highlights the possibility and advantage of working with a small number of human annotated labels, along with sparse, noisy but automatically acquired labels. The paper "*Automated methods to examine nonverbal synchrony in dyads*" reviews some of the theoretical and methodological challenges in studying interpersonal synchrony and proposes alternatives using automated computer vision techniques. The authors emphasize that the work can be a useful starting point for non-technical scholars looking to simplify their methods for examining synchrony. "*Multimodal sentiment and personality perception under speech: A comparison of transformer-based architectures*" compares distinct transformer architectures in sentiment estimation and personality perception, focusing on the attractive properties of the linear versions of transformers, and the use of appearance-invariant features with the objective of mitigating some appearance-related biases. Results suggest that each task may be better captured/modeled by a particular method. In "*Cracking the code of live human social interactions in autism: A review of the eye-tracking literature*", authors review the eye-tracking literature with a focus on characterizing the experimental paradigms and procedures to understand the scope and limitations in the context of human social interactions in autism. They discuss the theoretical implications of their findings and provide recommendations for future work, which will be essential to understand whether and how fundamental difficulties in perceiving and processing information about eye gaze cues interfere with social communication skills in autism. Last but not least, "*Covfee: an extensible web framework for continuous-time annotation of human behavior*" presents Covfee, an open source web annotation framework with crowd-sourcing support. Authors present results from case studies of continuous annotation of body poses (keypoints) and speaking (action) on an in-the-wild social interaction dataset. Advantages and disadvantages of applying continuous annotation to human behavior datasets are presented and discussed, in addition of general recommendations and other potential use cases.

Finally, we thank the program committee for their detailed reviews and support: Albert Ali Salah, Indu Bodala, Mohamed Chetouani, Enric Corona, Serkan Demirci, Ziya Erkoç, Ken Fujiwara, Rachael Jack, David Gallardo-Pujol, Uğur Güdükbay, Georgina Guilera, Hugo Jair Escalante, Isinsu Katircioglu, Taras Kucherenko, David Leiva, András Lőrincz, Catherine Pelachaud, David Saeteros, Hanan Salam, K. Suzanne Scherf, Sinan Sonlu, Vitomir Struc, Nguyen Tan Viet Tuyen, and Pieter Wolfert. We also thank all researchers that expressed their interest in contributing to this volume. We expect this volume to serve as a valuable reference and inspire future progress in social behavior understanding.