The 4th Workshop on Modeling Socio-Emotional and Cognitive Processes from Multimodal Data In-the-Wild (MSECP-Wild)

Dudzik, Bernd; Küster, Dennis; St-Onge, David; Putze, Felix

DOI
10.1145/3536221.3564029

Publication date
2022

Document Version
Final published version

Published in
ICMI 2022 - Proceedings of the 2022 International Conference on Multimodal Interaction

The 4th Workshop on Modeling Socio-Emotional and Cognitive Processes from Multimodal Data In-the-Wild (MSECP-Wild)

Bernd Dudzik
Delft University of Technology
Delft, The Netherlands
B.J.W.Dudzik@tudelft.nl

Dennis Küster
University of Bremen
Bremen, Germany
dennis.kuester@uni-bremen.de

David St-Onge
École de technologie supérieure
Montreal, Canada
david.st-onge@etsmtl.ca

Felix Putze
University of Bremen
Bremen, Germany
felix.putze@uni-bremen.de

ABSTRACT

The ability to automatically infer relevant aspects of human users' thoughts and feelings is crucial for technologies to adapt their behaviors in complex interactions intelligently (e.g., social robots or tutoring systems). Research on multimodal analysis has demonstrated the potential of technology to provide such estimates for a broad range of internal states and processes. However, constructing robust enough approaches for deployment in real-world applications remains an open problem. The MSECP-Wild workshop series serves as a multidisciplinary forum to present and discuss research addressing this challenge. This 4^{th} iteration focuses on addressing varying contextual conditions (e.g., throughout an interaction or across different situations and environments) in intelligent systems as a crucial barrier for more valid real-world predictions and actions. Submissions to the workshop span efforts relevant to multimodal data collection and context-sensitive modeling. These works provide important impulses for discussions of the state-of-the-art and opportunities for future research on these subjects.

CCS CONCEPTS

• **Human-centered computing → Empirical studies in ubiquitous and mobile computing**.

KEYWORDS

User-Modeling, Multimodal Data, Affective Computing, Social Signal Processing, Ubiquitous Computing, Context-awareness

ACM Reference Format:
Bernd Dudzik, Dennis Küster, David St-Onge, and Felix Putze. 2022. The 4th Workshop on Modeling Socio-Emotional and Cognitive Processes from Multimodal Data In-the-Wild (MSECP-Wild). In *INTERNATIONAL CONFERENCE ON MULTIMODAL INTERACTION (ICMI '22), November 7–11, 2022, Bengaluru, India.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3536221.3564029

ICMI '22, November 7–11, 2022, Bengaluru, India

ACM ISBN 978-1-4503-9390-4/22/11.
https://doi.org/10.1145/3536221.3564029

1 INTRODUCTION

Modern intelligent systems are expected to support human actions and decisions in complex task environments. Importantly, this increasingly includes domains where performance relates to human psycho-social needs. Examples range from personalized entertainment systems to intelligent tutoring applications to social robots in elderly care. For such systems to display adaptive behavior in real-world environments (i.e., *In the Wild*), they need to understand how their users think and feel.

Research on multimodal analysis of behavioral and physiological data has demonstrated the potential to provide estimates of internal states and processes, e.g., a person's attentional engagement. However, despite substantial technological advances, important theoretical and practical challenges for reliably detecting internal states remain unresolved, hindering progress towards more successful real-world applications. Such issues span low-level processing and integration of noisy data streams, conceptual pitfalls, data collection, and pressing ethical questions about what intelligent systems should (not) do. Together these issues issues highlight the need for interdisciplinary collaboration [1].

The present workshop brings academics and practitioners together to discuss recent contributions relating to overcoming these challenges. Like previous iterations [9–11], it is a concerted effort to stimulate joint research projects, an exchange of methods, and a critical discussion of current and future efforts. In the following, we provide a brief overview of the submissions.

2 WORKSHOP CONTENT

A key challenge for robust predictions of internal states in real-world applications relates to the broad range of changing contextual conditions under which interactions with such systems occur [7]. However, despite the benefits of addressing context for predictions in the wild [4], existing technological research has only scarcely addressed this issue. Instead, efforts have largely focused on the context-free analysis of human behavioral signals. Integrating both stable and dynamic contexts into multimodal modeling remains an open problem. Some key barriers to context-sensitive solutions are that it is (1) unclear what constitutes relevant contextual information, i.e., information that is *effective* for improving multimodal predictions [6], as well as (2) how to *feasibly* obtain and incorporate this information into technologies - while (3) avoiding the potential *combinatorial explosion* of possible contexts and its impact on the

amount of required modeling data. Contributions to our workshop have aimed to develop strategies for addressing key contextual factors in multimodal modeling across a broad scope of approaches:

Contextual Modulation of Affect: Comparing Humans and Deep Neural Networks. Shin et al. [12] explore the correspondence between existing context-sensitive deep learning architectures for vision-based automatic affect recognition and human emotion perception. Their findings indicate limitations in existing models to process context in a human-like capacity.

How can Interaction Data be Contextualized with Mobile Sensing to Enhance Learning Engagement Assessment in Distance Learning? Ciordas-Hertel et al. [2] describe a technological architecture to collect multimodal data contextualizing students' interactions with a system for remote learning. Furthermore, it presents findings from a user study with promising results for future research on an adaptive application.

Exploring the Benefits of Spatialized Multimodal Psychophysiological Insights for User Experience Research. Simard et al. [13] present an industry contribution describing a novel data collection platform for psychophysiological research in the wild. The platform facilitates capturing physiological (e.g., EEG and EDA) in combination with indoor location data. The article presents preliminary findings from data collection in two public events.

Improving Supervised Learning in Conversational Analysis through Reusing Preprocessing Data as Auxiliary Supervisors. Kim et al. [8] present a multi-task learning framework for audiovisual affect recognition in conversations with predictions at the speaker-turn level as a primary task. Notably, their approach employs emotion labels for past and future speaker-turns as auxiliary tasks to provide temporal context for predictions.

Predicting evaluations of entrepreneurial pitches based on multimodal nonverbal behavioral cues and self-reported characteristics. Stoitsas et al. [14] outline a multimodal approach for predicting investor's assessment of entrepreneurial pitches by fusing user profiles with the non-verbal behavior of both parties. Consistent with prior work by [5], the behavior of an individual's interaction partner can produce vital contextual information for modeling their cognitions and actions in social situations.

Investigating Transformer Encoders and Fusion Strategies for Speech Emotion Recognition in Emergency Call Center Conversations. . Deschamps-Berger et al. [3] investigate the use of pre-trained and fine-tuned Transformer models for audio and text modalities for emotion recognition. It provides a use-case of how to apply pre-trained machine learning architectures to deal with limited data available in specific contexts.

3 CONCLUSIONS

MSECP-Wild hosts a variety of submissions from academia and industry relevant to the challenge of context for applications in the wild. It covers contextual data collection and sparsity issues, as well as context-sensitive modeling and evaluation. Apart from presentations, we have invited speakers offering expertise on different forms of context and its integration into intelligent systems. Together, this provides a platform for interdisciplinary exchange about addressing context in the wild in future research.

ACKNOWLEDGMENTS

This research was (partially) funded by the Hybrid Intelligence Center, a 10-year program funded by the Dutch Ministry of Education, Culture, and Science through the Netherlands Organisation for Scientific Research, https://hybrid-intelligence-centre.nl, grant number 024.004.022.

REFERENCES

[1] Patrícia Alves-Oliveira, Dennis Küster, Arvid Kappas, and Ana Paiva. 2016. Psychological science in hri: Striving for a more integrated field of research. In *2016 AAAI Fall Symposium Series*.

[2] George-Petru Ciordas-Hertel, Daniel Biedermann, Marc Winter, Julia Mordel, and Hendrik Drachsler. 2022. How can Interaction Data be Contextualized with Mobile Sensing to Enhance Learning Engagement Assessment in Distance Learning?. In *INTERNATIONAL CONFERENCE ON MULTIMODAL INTERACTION (ICMI '22 Companion)*. ACM.

[3] Theo Deschamps-Berger, Lori Lamel, and Laurence Devillers. 2022. Investigating Transformer Encoders and Fusion Strategies for Speech Emotion Recognition in Emergency Call Center Conversations. In *INTERNATIONAL CONFERENCE ON MULTIMODAL INTERACTION (ICMI '22 Companion)*. ACM.

[4] Bernd Dudzik, Joost Broekens, Mark Neerincx, and Hayley Hung. 2020. Exploring Personal Memories and Video Content as Context for Facial Behavior in Predictions of Video-Induced Emotions. *Proceedings of the 2020 International Conference on Multimodal Interaction* 10, 153–162. Issue 20. https://doi.org/10.1145/3382507.3418814

[5] Bernd Dudzik, Simon Columbus, Tiffany Matej Hrkalovic, Daniel Balliet, and Hayley Hung. 2021. Recognizing Perceived Interdependence in Face-to-Face Negotiations through Multimodal Analysis of Nonverbal Behavior. *Proceedings of the 2021 International Conference on Multimodal Interaction*, 121–130. Issue 1. https://doi.org/10.1145/3462244.3479935

[6] Bernd Dudzik, Michel-Pierre Jansen, Franziska Burger, Frank Kaptein, Joost Broekens, Dirk K.J. Heylen, Hayley Hung, Mark A. Neerincx, and Khiet P. Truong. 2019. Context in Human Emotion Perception for Automatic Affect Detection: A Survey of Audiovisual Databases. In *2019 8th International Conference on Affective Computing and Intelligent Interaction (ACII)*. IEEE, 206–212. https://doi.org/10.1109/ACII.2019.8925446

[7] Zakia Hammal and Merlin Teodosia Suarez. 2015. Towards context based affective computing introduction to the third international CBAR 2015 workshop. In *2015 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG)*. IEEE, 1–2. https://doi.org/10.1109/FG.2015.7284841

[8] Joshua Y. Kim, Tongliang Liu, and Kalina Yacef. 2022. Improving Supervised Learning in Conversational Analysis through Reusing Preprocessing Data as Auxiliary Supervisors. In *INTERNATIONAL CONFERENCE ON MULTIMODAL INTERACTION (ICMI '22 Companion)*. ACM.

[9] Dennis Küster, Felix Putze, Patrícia Alves-Oliveira, Maike Paetzel, and Tanja Schultz. 2020. Modeling Socio-Emotional and Cognitive Processes from Multimodal Data in the Wild. *ICMI 2020 - Proceedings of the 2020 International Conference on Multimodal Interaction* (2020), 883–885. https://doi.org/10.1145/3382507.3420053

[10] Dennis Küster, Felix Putze, David St-Onge, Pascal E. Fortin, Nerea Urrestilla, and Tanja Schultz. 2021. 3rd Workshop on Modeling Socio-Emotional and Cognitive Processes from Multimodal Data in the Wild. *ICMI 2021 - Proceedings of the 2021 International Conference on Multimodal Interaction* (2021), 860–861. https://doi.org/10.1145/3462244.3480978

[11] Felix Putze, Enkelejda Kasneci, Jutta Hild, Erin Solovey, Akane Sano, and Tanja Schultz. 2018. Modeling cognitive processes from multimodal signals. *ICMI 2018 - Proceedings of the 2018 International Conference on Multimodal Interaction* (2018), 663. https://doi.org/10.1145/3242969.3265861

[12] Soomin Shin, Doo Yon Kim, and Christian Wallraven. 2022. Contextual Modulation of Affect: Comparing Humans and Deep Neural Networks. In *INTERNATIONAL CONFERENCE ON MULTIMODAL INTERACTION (ICMI '22 Companion)*. ACM.

[13] Frederic Simard, Tony Aumont, Sayeed A. D. Kizuk, and Pascal E. Fortin. 2022. Exploring the Benefits of Spatialized Multimodal Psychophysiological Insights for User Experience Research. In *INTERNATIONAL CONFERENCE ON MULTIMODAL INTERACTION (ICMI '22 Companion)*. ACM.

[14] Kostas Stoitsas, Itır Önal Ertuğrul, Werner Liebregts, and Merel M. Jung. 2022. Predicting evaluations of entrepreneurial pitches based on multimodal nonverbal behavioral cues and self-reported characteristics. In *INTERNATIONAL CONFERENCE ON MULTIMODAL INTERACTION (ICMI '22 Companion)*. ACM.