Social Signals and Multimedia

Past, Present, Future

Hung, Hayley; Gurrin, Cathal; Larson, Martha; Gunes, Hatice; Ringeval, Fabien ; Andre, Elisabeth ; Morency, Louis-Philippe

DOI
10.1145/3474085.3480024

Publication date
2021

Document Version
Final published version

Published in
MM 2021

Citation (APA)
Hung, H., Gurrin, C., Larson, M., Gunes, H., Ringeval, F., Andre, E., & Morency, L-P. (2021). Social Signals and Multimedia: Past, Present, Future. In *MM 2021 : Proceedings of the 29th ACM International Conference on Multimedia* (pp. 4610-4612). Association for Computing Machinery (ACM). https://doi.org/10.1145/3474085.3480024

Social Signals and Multimedia: Past, Present, Future

Hayley Hung
Delft University of Technology
The Netherlands
h.hung@tudelft.nl

Cathal Gurrin
Dublin City University
Ireland

Martha Larson
Radboud University
The Netherlands

Hatice Gunes
University of Cambridge
United Kingdom

Fabien Ringeval
Université Grenoble Alpes
France

Elisabeth Andre
University of Augsburg
Germany

Louis-Philippe Morency
Carnegie Mellon University
USA

ABSTRACT

The rising popularity of Artificial Intelligence (AI) has brought considerable public interest as well faster and more direct transfer of research ideas into practice. One of the aspects of AI that still trails behind considerably is the role of machines in interpreting, enhancing, modeling, generating, and influencing social behavior. Such behavior is captured as social signals, usually by sensors recording multiple modalities, making it classic multimedia data. Such behavior can also be generated by an AI system when interacting with humans. Using AI techniques in combination with multimedia data can be used to pursue multiple goals, two of which are highlighted here. First, supporting people during social interactions and helping them to fulfil their social needs either actively or passively. Second, improving our understanding of how people collaborate, build relationships, and process self identity. Despite the rise of fields such as Social Signal Processing, a similar panel organised at ACM Multimedia 2014, and an area on social and emotional signals at the ACM MM since 2014, we argue that we have yet to truly fulfil the potential of the combining social signals and multimedia. This panel asks where we have come far enough and what remaining challenges there are in light of recent global events.

CCS CONCEPTS

• **Applied computing** → **Psychology**; • **Human-centered computing** → **Collaborative and social computing theory, concepts and paradigms**; • **Computing methodologies** → **Artificial intelligence**; • **Information systems** → **Multimedia information systems**.

KEYWORDS

multimedia, human social behavior, artificial intelligence, multimodal machine learning, social signal processing

MM '21, October 20–24, 2021, Virtual Event, China

ACM ISBN 978-1-4503-8651-7/21/10.
https://doi.org/10.1145/3474085.3480024

ACM Reference Format:
Hayley Hung, Cathal Gurrin, Martha Larson, Hatice Gunes, Fabien Ringeval, Elisabeth Andre, and Louis-Philippe Morency. 2021. Social Signals and Multimedia: Past, Present, Future. In *Proceedings of the 29th ACM International Conference on Multimedia (MM '21), October 20–24, 2021, Virtual Event, China.* ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/3474085.3480024

1 INTRODUCTION

This panel aims to address the bottlenecks and opportunities facing the Multimedia Computing community and researchers in adjacent fields as they push forward the frontier of AI related to humans and machines. Specifically, we are interested in the following:

(1) AI that is used as the basis for applications that support human social interaction;
(2) AI that allows us to gain a fundamental understanding of human social behavior, even to the extent of developing novel social science theories;
(3) Social interaction support interfaces ranging from passive interfaces to AI being active social participants.

With the global pandemic and social distancing, applications that can help improve our social well-being has gone from a 'nice to have' to a key component for combating increasing mental health problems related to social isolation. Although we have scrambled to develop new ways to socialize virtually, we have lost the possibility to have in-person, serendipitous social experiences. Despite attempts that have been made to recreate informal and chance encounters virtually, much of the social bonding that comes from in person interaction has also been lost. It has become clear that looking after social well-being is a vital part of preventing more chronic mental health issues such as loneliness and depression. Despite the growth in AI in recent years, we have to ask ourselves whether we were as ready as we could have been to produce technologies to help people to fulfil their social needs. Much of current AI research still focuses less on understanding 'social' phenomena and more on understanding more basic primitive machine perception behavioral concepts such as human action recognition. When behavior prediction or understanding is attempted, the modeling of the cognitive motivations behind an individual's behavior are either not considered at all, or they are hypothesised posthoc. While there is an inherent comfort to solving tasks that involve objective ground

truth, the biggest challenge is how to learn from real world data occurring in real-world situations where real-life social relationships are at stake.

We take the position that Social Multimedia Computing should be there to help to enhance, inform, or influence our social interactions for more effective and satisfying social interaction experiences. However, training models to a sufficient level of robustness requires significant amounts of data that cannot be collected using the more popular recent approaches involving trawling the internet for online 'found data'. Real data of real relationships is often sparse, messy, complicated, and inherently more private. This situation puts us into a conundrum as to how to train new models for such machine understanding tasks. New paradigms of data collection, sharing, and processing are likely to be needed.

2 BACKGROUND

Aside from the global events that have taken place in recent years, the idea of integrating Social Signals into Multimedia Computing is nothing new. However, one can already observe that a potential bias exists in our understanding of what Social Signals are. A quick Google search for 'Social Signals' will bring up results predominantly about social media signals and marketing strategies. This was also the case back in 2014 when a related panel "Emotional and Social Signals: Where art Thou?" was organized [2]. Given the importance of search engines and data on the internet for helping us to understand and study the world through the eyes of machines considers this seems an interesting disparity. However, the term originated from Social Science and referred to signals generated from a constellation of human behavioural indicators or cues [5] coming from body language.

The magnitude of our lack of understanding of the perils of found data is beginning to show. See the extensive discussion by Prabhu et al. [4] about the large scale misuse of human data for training computer vision models. Increased concerns for responsible practices in AI have led to more debates in the public arena about the sourcing of human data. This also makes the sourcing of human social interaction data even more sensitive. This caution has also been observed in the industrial market place such as the decision by Apple and Google during their recent collaboration for smart mobile based contact tracing [1]. Design choices were made to ensure that while bluetooth proximity information could be shared, location information would not be.

If found data from the internet cannot be so easily shared openly, researchers fall back on more prohibitive sharing licences that restrict access to research institutions only. These are more difficult to handle administratively as manual intervention is required to ensure that data applicants are reputable and misuse of data is policed to some degree. Funds to continue looking after datasets that are available only under restrictive licences are usually not available beyond the life of an associated project. Given the manual effort involved, sharing data of social signals is likely to have much less impact than the publicly open datasets containing millions of examples that can easily be trawled online. When considering the gathering of highly personal human behavioral data in long term social settings, the situation is compounded. In those particular settings obtaining the ground truth is an additional complexity calling on innovative approaches for experience sampling [3].

Aside from the concerns mentioned above, some forms of unwanted bias can still exist in the resultant system. The social science theories we base our systems on are in themselves evolving. The intervention of the researcher is vitally important. Leading by in example in responsible system design helps to ensure that realistic expectations about the technology are communicated. Education to the general public as well as the next generation of practitioners and researchers to be responsible and mindful of how such technologies could be used is also crucial. We must also not overlook recent developments in AI related to privacy preserving methodologies [2]

There may be an opportunity to fuse some of the concerns and opportunities mentioned above. One such possibility is to consider an approach that frames the relationship between the humans and machines within the ecosystem of a more collaborative, adaptive, responsible, and ethical framework - a so-called Hybrid Intelligence approach [1]. A successful implementation of this is still an open question but could include paradigms involving learning and adapting from limited data and more human intervention. Such topics may well be more difficult for industry to take a lead on and perhaps calls for more leadership and initiative from research institutions.

3 PANEL TOPICS

The panel will consider the past, present, and future of AI research at the intersection of social signals and multimedia. Questions that are of specific interest are the following:

- What are the most important achievements in the study of social signal processing in the fields of multimedia and adjacent fields? What are historically the missed opportunities?
- What are the relevance of social signals given the current popularity of AI research? Where can the Multimedia Community take the lead? Where might it be lagging behind?
- What are the challenges and opportunities of moving from generalized to (life-long) personalized models?
- How can we ensure that the research community is invested appropriate effort into creating and annotating the necessary data sets?
- How do we design such data sets to ensure that the research or applications using them is fair and avoids bias? How do we educate users about potential (unintended) misuses?
- How can we promote a responsible understanding of research outcomes involving AI and social signals. This also includes promoting the best grounding to our knowledge based on social science theories?
- What must we do to ensure that ethics and data privacy are appropriately addressed? How can we make clear to the community that multimedia social signals research done right does not need to violate privacy (privacy preserving algorithms, best practices in data collection)?

ACKNOWLEDGMENTS

H. Hung is partially funded by the Netherlands Organization for Scientific Research (NWO) under the MINGLE project number

[1] https://covid19.apple.com/contacttracing

[2] https://ppml-workshop.github.io/

639.022.606. H. Gunes is supported by the EPSRC project ARoEQ under grant ref. EP/R030782/1, and partially by the European Union's Horizon 2020 Research and Innovation Programme Project WorkingAge, under grant agreement No. 826232.

REFERENCES

[1] Zeynep Akata, Dan Balliet, Maarten de Rijke, Frank Dignum, Virginia Dignum, Guszti Eiben, Antske Fokkens, Davide Grossi, Koen Hindriks, Holger Hoos, et al. 2020. A Research Agenda for Hybrid Intelligence: Augmenting Human Intellect With Collaborative, Adaptive, Responsible, and Explainable Artificial Intelligence. *Computer* 53, 8 (2020), 18–28.

[2] Hatice Gunes and Hayley Hung. 2015. Emotional and Social Signals: A Neglected Frontier in Multimedia Computing? *IEEE MultiMedia* 22, 2 (2015), 76–85.

[3] Abhinav Mehrotra, Jo Vermeulen, Veljko Pejovic, and Mirco Musolesi. 2015. Ask, but don't interrupt: the case for interruptibility-aware mobile experience sampling. In *Adjunct Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2015 ACM International Symposium on Wearable Computers*. ACM, Osaka, Japan, 723–732.

[4] Vinay Uday Prabhu and Abeba Birhane. 2020. Large image datasets: A pyrrhic win for computer vision? arXiv:2006.16923 [cs.CY]

[5] Alessandro Vinciarelli, Maja Pantic, and Hervé Bourlard. 2009. Social signal processing: Survey of an emerging domain. *Image and vision computing* 27, 12 (2009), 1743–1759.