

Socially Intelligent Affective AI

Aarti Malhotra

David R. Cheriton School of Computer Science
University of Waterloo
Waterloo, Canada
aarti.malhotra@uwaterloo.ca

Abstract

Artificial Intelligence has aimed to give the systems or agents, the ability to learn, perceive, recognize, plan, reason and act. Affective Computing has brought into focus the importance of giving AI systems, the capability to perceive, detect, utilize and generate emotion, affect, sentiment or feelings. To have a meaningful human-computer interaction, we need to design and develop a more socially intelligent and affective AI. My doctoral research goal is to delve deeper into some of these aspects, firstly by surveying computational models implemented in AI that uses emotion in decision-making or behaviour; secondly, by creating new model to predict social event context and affect in group videos; thirdly, to predict the social identities in visual scenes; and lastly to combine information about context, identities, behaviour and emotion in a social interaction scene to predict social incoherence and to recommend appropriate behaviour.

Introduction

“All the world’s a stage, And all the men and women merely players; They have their exits and their entrances, And one man in his time plays many parts”

William Shakespeare

The above words by William Shakespeare beautifully describes the social reality and how we, as humans, play various roles and do our part. Emotion, affect, sentiment, feeling is a core part of our being and they guide our decisions and behaviour. Understanding the context, identities and behaviours in a situation gives us a holistic view and helps us in our day-to-day interactions. During my PhD research, it was inspiring to tap into some aspects of our underlying affective processing and social intelligence, and aiming to give AI such capabilities. A comprehensive computational model of emotions in social interaction is *Affect Control Theory (ACT)* (Hoey, Schröder, and Alhothali 2016), includes components for context, identities, normative behaviours, emotions, culturally shared out-of-context sentiments and an in-context impressions of an interaction. A recent extension of ACT, called *Bayesian Affect Control Theory (BayesACT)* (Hoey, MacKinnon, and Schröder 2021), puts uncertainty at

center stage by explicitly modeling variance in emotional sentiments about identities and contexts. Although ACT is based on human empirical data from various cultural groups and is quite promising, it is yet to be adopted in AI applications. In this thesis, I survey other existing computational models of emotion that incorporate emotions in behaviour. I perform simulations of ACT in real-world scenarios. A computer vision system currently does not exist for ACT models that can take input as images or videos, detect context, identities, behaviour and predict emotional incoherence and recommend behaviour to better align agent with human in an interaction. This thesis builds predictive models of social event context, social identities and an ACT based network to predict incoherence, along with behaviour recommendation.

Emotion in Social Decision-Making and Behaviour

Much of affective computing field has dealt with emotion recognition and emotion generation, but not much focus has been given to incorporate emotions into decision-making and behaviour in AI agents to make it more socially aligned with humans. Modelling such a system could give insights about human affective process, and would be a step towards systems being more human-like and less robotic.

RQ1: What are the existing computational models of emotion which have both an AI implementation in a system that considers emotion to guide decision-making and behaviour?

Progress:

- *Current work:* A systematic literature review is in progress and a search is done on Scopus, IEEE, ACM
- *Plan:* Complete screening and review in Fall 2021. Detailed analysis and writing of the paper in Winter 2022.
- *Target venue:* TAFFC Journal

ACT Simulations

For long-term adoption of ACT in human-agent social interaction, there is a need to design, simulate and validate in an affective social agent in real-world settings.

RQ2: How can an affective AI agent be designed and simulated in a realistic scenario using ACT/BayesACT?

Progress:

- *Completed work:* a. Paper on ‘Emotions in Socio-cultural Interactive AI Agents’ (Malhotra and Hoey 2021) accepted at ACII workshop on ‘Functions of emotions for socially interactive agents’. b. Paper on ACT simulation in a brain-simulation framework called ‘Nengo’, depicting how perceived affective social interaction may guide behaviour, is accepted at ICCM (Malhotra, Stewart, and Hoey 2020). c. Manuscript on ACT simulation titled ‘User perspectives on emotionally aligned social robots for older adults and persons with dementia’ (Dosso, Bandari, Malhotra, Hoey, Michaud, Prescott, Robillard) submitted to ‘The Journal of Rehabilitation and Assistive Technologies Engineering’
- *Current work:* Prisoner’s dilemma game play using ACT and BayesACT bot with human player
- *Plan:* To complete game experiments and user survey
- *Target venue:* AAMAS or ICMI

Social Context and Affect Prediction

Understanding social context and affect plays a key role in our day-to-day interactions and enables us to understand better the situation, our own self, and others. Context in humans has various facets such as exteroceptive world related location, event, people, socio-cultural, political or interoception such as body regulation and many more. Classifying a social event and associated perceived affect in a visual scene involving humans, is a step towards providing better visual understanding capability to socially interactive AI agents.

RQ3: How well can a computer vision system predict social events and emotions in a visual scene? Does affect information help improve the classification of social events?

Progress:

- *Current work:* Novel spatio-temporal transformer network for multi-task learning, that uses Video level Group Affect (VGAF) dataset (Sharma, Dhall, and Cai 2021) is developed, improving state of the art for affect prediction and providing baseline for social event prediction, with improvements over CNNs and other transformers.
- *Plan:* To perform experiments on another dataset.
- *Target venue:* ICPR or ECCV or AAAI

Social Identities Prediction

As AI agents become a part of our day-to-day life, successful human-machine interaction becomes an essential part of the experience. Understanding the nuances of human social interaction remains a challenging area of research, but there is a growing consensus that emotional identity, or what the social face a person presents in a given context, is a critical aspect. Therefore, understanding the identities displayed by humans, and the identity of agents in the social context, is a crucial skill for a socially interactive agent.

RQ4: How well can a computer vision system predict social identities and/or relationships in a visual scene?

Progress:

- *Completed Work:* Paper on ‘The Role of Identities in Socially Interactive Affective Robots’ (Malhotra and Hoey 2021) accepted at HRI workshop on Robo-Identities
- *Current Work:* Survey of suitable datasets, like (Li et al. 2017), having identities annotations, is in progress
- *Plan:* To build an identities and/or relationship classifier.
- *Target venue:* CVPR or ECCV or AAAI

Social Incoherence Prediction and Behaviour Recommendation

Social norms emerge from the efforts of society members to create stability and coherence. If some group members do not adhere to the norms, then it is highly likely that there will be a conflicting situation and cause social incoherence.

RQ5: How can information about context, identities and behaviour in a visual scene of social interaction be utilized to predict emotional incoherence level and recommend a suitable behaviour to minimize the conflict?

Progress:

- *Current Work:* An initial version of inference network is built using pre-trained models for place context, identities, and behaviour. Work is in progress to have a better performing identity and behaviour classifier. Datasets to validate may be TV Human interactions, MovieGraphs.
- *Plan:* To complete the network and validate.
- *Target venue:* ICCV or Pattern Recognition Letters or Computer Vision and Image Understanding

Future Work

In future, any of the classifiers can take into consideration multi-modal inputs of speech and text to be more informative of the scene. Also, some specific application domain modeling such as healthcare, education, etc., can pave the way for practical usages in virtual agents or robots.

References

- Hoey, J.; MacKinnon, N. J.; and Schröder, T. 2021. Denotative and Connotative Management of Uncertainty: A Computational Dual-Process Model. *Judgement and Decision Making*, 16(2).
- Hoey, J.; Schröder, T.; and Alhothali, A. 2016. Affect Control Processes: Intelligent Affective Interaction using a Partially Observable Markov Decision Process. *Artificial Intelligence*, 230: 134–172.
- Li, J.; Wong, Y.; Zhao, Q.; and Kankanhalli, M. S. 2017. Dual-glance model for deciphering social relationships. In *Proceedings of the IEEE International Conference on Computer Vision*, 2650–2659.
- Malhotra, A.; Stewart, T. C.; and Hoey, J. 2020. A Biologically-Inspired Neural Implementation of Affect Control Theory. *International Conference on Cognitive Modelling, Toronto*.
- Sharma, G.; Dhall, A.; and Cai, J. 2021. Audio-visual Automatic Group Affect Analysis. *IEEE Transactions on Affective Computing*, (01): 1–1.