An Ethically-Guided Domain-Independent Model of Computational Emotions

Suman Ojha

Faculty of Engineering and Information Technology
University of Technology Sydney

This dissertation is submitted for the degree of
Doctor of Philosophy

Supervisor:
Dist. Prof. Mary-Anne Williams

Co-supervisor:
Dr. Jonathan Vitale

February 2020

This research is dedicated to my father, Dr. Narayan Ojha, who wanted to see me as a medical doctor like him. Although, I took a path of engineering instead of medicine, I have marched on my way to become a 'doctor' – though not in medicine but in philosophy of Computer Science.

Certificate of Original Authorship

I, Suman Ojha, declare that this thesis is submitted in fulfilment of the requirements for the award of Doctor of Philosophy, in the School of Computer Science, Faculty of Engineering and Information Technology at the University of Technology Sydney.

This thesis is wholly my own work unless otherwise referenced or acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis.

This document has not been submitted for qualifications at any other academic institution.

This research is supported by the Australian Government Research Training Program.

Signature of Student:

Production Note:
Signature removed prior to publication.

Suman Ojha
24 February 2020

Acknowledgements

My research project would not have been successful without the help and support of a number of people. I hereby take an opportunity to personally acknowledge everyone who walked with me in their own ways along the journey of this PhD research. I would like to heartily thank the following people.

- My PhD supervisor, Distinguished Professor Mary-Anne Williams, for being an exemplary advisor. She allowed me to pave my own path for the completion of this journey and supported along the way in every possible manner. She also taught me that paper rejections are the norms as an early career researcher and one should take this as an opportunity to refine one's work and become a better researcher.
- My co-supervisor, Dr. Jonathan Vitale, for being more of a colleague than just a supervisor and helping me in uncountable ways.
- My wife, Asmita Thapa, who took my research as seriously as I did, and went through everything to keep me away from the stress of my studies.
- Dr. Syed Ali Raza and Dr. Richard Billingsley for providing me useful concepts of machine learning approaches for the completion of my research project.
- My family and friends who always treated me as a person rather than just a nerdy research student and never failed to make me realise that I was alive.
- My colleagues at the Innovation and Enterprise Research Lab (The Magic Lab) who helped to improve my research by attending my presentations and providing feedback on the drafts of my papers.
- All the anonymous participants of my research surveys who contributed their valuable data for the evaluation of my computational model of emotion.
- All the anonymous reviewers of my conference and journal papers who provided valuable comments and suggestions for the improvement of my research.

I consider you all as a part of my research journey.

Author's Core Research Contributions

Ojha, S., Gudi, S. L. K. C., Vitale, J., Williams, M.-A., and Johnston, B. (2017a). I remember what you did: A bheavioural guide-robot. In *International Conference on Robot Intelligence Technology and Applications (RiTA)*.

Ojha, S., Vitale, J., Ali Raza, S., Billingsley, R., and Williams, M.-A. (2018a). Implementing the dynamic role of mood and personality in emotion processing of cognitive agents. In *Annual Conference on Advances in Cognitive Systems (ACS)*.

Ojha, S., Vitale, J., Ali Raza, S., Billingsley, R., and Williams, M.-A. (2019). Integrating mood and personality with agent emotions. In *International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*.

Ojha, S., Vitale, J., and Williams, M.-A. (2017b). A domain-independent approach of cognitive appraisal augmented by higher cognitive layer of ethical reasoning. In *Annual Meeting of the Cognitive Science Society (CogSci)*.

Ojha, S. and Williams, M.-A. (2016). Ethically-guided emotional responses for social robots: Should i be angry? In *International Conference on Social Robotics (ICSR)*, pages 233–242. Springer.

Ojha, S. and Williams, M.-A. (2017). Emotional appraisal: A computational perspective. In *Annual Conference on Advances in Cognitive Systems (ACS)*.

Ojha, S., Williams, M.-A., and Johnston, B. (2018b). The essence of ethical reasoning in robot-emotion processing. *International Journal of Social Robotics (IJSR)*, 10:211–223.

Author's Additional Research Contributions

Gudi, S. L. K. C., Ojha, S., Alam, S., Johnston, B., and Williams, M.-A. (2017a). A proactive robot tutor based on emotional intelligence. In *International Conference on Robot Intelligence Technology and Applications (RiTA)*.

Gudi, S. L. K. C., Ojha, S., Clark, J., Johnston, B., and Williams, M.-A. (2017b). Fog robotics: An introduction. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (iROS)*.

Gudi, S. L. K. C., Ojha, S., Johnston, B., Clark, J., and Williams, M.-A. (2018). Fog robotics for efficient, fluent and robust human-robot interaction. In *IEEE 17th International Symposium on Network Computing and Applications (NCA)*.

Herse, S., Vitale, J., Ebrahimian, D., Tonkin, M., Ojha, S., Sidra, S., Johnston, B., Phillips, S., Gudi, S. L. K. C., Clark, J., et al. (2018a). Bon appetit! robot persuasion for food recommendation. In *Companion of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*, pages 125–126. ACM.

Herse, S., Vitale, J., Tonkin, M., Ebrahimian, D., Ojha, S., Johnston, B., Judge, W., and Williams, M.-A. (2018b). Do you trust me, blindly? factors influencing trust towards a robot recommender system. In *International Symposium on Robot and Human Interactive Communication (RO-MAN)*.

Tonkin, M., Vitale, J., Ojha, S., Clark, J., Pfeiffer, S., Judge, W., Wang, X., and Williams, M.-A. (2017a). Embodiment, privacy and social robots: May i remember you? In *International Conference on Social Robotics (ICSR)*, pages 506–515. Springer.

Tonkin, M., Vitale, J., Ojha, S., Williams, M.-A., Fuller, P., Judge, W., and Wang, X. a. (2017b). Would you like to sample? Robot engagement in a shopping centre. In *International Symposium on Robot and Human Interactive Communication (RO-MAN)*. IEEE.

Vitale, J., Tonkin, M., Herse, S., Ojha, S., Clark, J., Williams, M.-A., Wang, X., and Judge, W. (2018). Be more transparent and users will like you: A robot privacy and user experience design experiment. In *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*, pages 379–387. ACM.

Vitale, J., Tonkin, M., Ojha, S., Williams, M.-A., Wang, X., and Judge, W. (2017). Privacy by design in machine learning data collection: A user experience experimentation. In *AAAI Spring Symposium Series*.

Table of Contents

List of figures

List of tables

Abstract

Advancement of artificial intelligence research has supported the development of intelligent autonomous agents. Such intelligent agents, like social robots, are already appearing in public places, homes and offices. Unlike the robots intended for use in factories for mechanical work, social robots should not only be proficient in capabilities such as vision and speech, but also be endowed with other human skills in order to facilitate a sound relationship with human counterparts.

Phenomena of emotions is a distinguishable human feature that plays a significant role in human social communication because ability to express emotions enhances the social exchange between two individuals. As such, artificial agents employed in social settings should also exhibit adequate emotional and behavioural abilities to be easily adopted by people.

A critical aspect to consider when developing models of artificial emotions for autonomous intelligent agents is the likely impact that the emotional interaction can have on the human counterparts. For example, an *emotional robot* that shows an angry expression along with a loud voice may scare a young child more than a *non-emotional robot* that only denies a request. Indeed, most modern societies consider a strong emotional reaction towards a young child to be unacceptable and even unethical.

How can a robot select a socially acceptable emotional state to express while interacting with people? I answer this question by providing an association between emotion theories and ethical theories – which has largely been ignored in the existing literature.

A regulatory mechanism for artificial agents inspired by ethical theories is a viable way to ensure that the emotional and behavioural responses of the agent are acceptable in a given social context. As such, an intelligent agent with emotion generation capability can establish social acceptance if its emotions are regulated by ethical reasoning mechanism.

In order to validate the above statement, in this work, I provide a novel computational model of emotion for artificial agents – EEGS (short name for **E**thical **E**motion **G**eneration **S**ystem) and evaluate it by comparing the emotional responses of the model with emotion data collected from human participants. Experimental results support that *ethical reasoning mechanism can indeed help an artificial agent to reach to a socially acceptable emotional state*.