Emotion-based Norm Identification

JIETING LUO (LUORA)¹, THOMAS STUDER¹, MEHDI DASTANI²

UNIVERSITY OF BERN

UTRECHT UNIVERSITY

Agents need to reason about emotions

 Designing artificial systems with emotions: it can help to get artefacts with a more human-like behavior thus having deeper and more meaningful humanmachine relationship.

Emotional expressions provide pertinent cues into the internal states of expressers. These cues play a critical role in social interaction by serving as communication devices that facilitate human-robot interaction.

Agents need to reason about emotions





Modeling Emotions

•OCC psychological model: emotions are structured based on **focus of attention**.



Modeling Emotions



Modeling Emotions

Definition 3.1 (Joy and Distress). Given agent i, an action a and a value v,

$$\operatorname{joy}_{i}(a, v) \stackrel{def}{=} B_{i}(v \wedge \langle -a \rangle \neg v) \wedge \operatorname{Val}_{i}(v),$$

distress_{i}(a, v) \stackrel{def}{=} B_{i}(\neg v \wedge \langle -a \rangle v) \wedge \operatorname{Val}_{i}(v).

Definition 3.2 (Hope and Fear). Given agent i, an action a and a value v,

hope_i(a, v)
$$\stackrel{def}{=} B_i \neg v \land \text{uncertain}_i(a, v) \land \text{Val}_i(v),$$

fear_i(a, v) $\stackrel{def}{=} B_i v \land \text{uncertain}_i(a, v) \land \text{Val}_i(v),$
where uncertain_i(a, v) = $\neg B_i \langle a \rangle v \land \neg B_i \langle a \rangle \neg v \land B_i \langle a \rangle \top.$

Modeling Inferential Process

 $\models B_i(\mathrm{joy}_j(a,v) \lor \mathrm{distress}_j(a,v) \lor \mathrm{hope}_j(a,v) \lor \mathrm{fear}_j(a,v)) \to B_i \operatorname{Val}_j(v).$



Inductive Logic Programming:

Awad, Edmond, Michael Anderson, Susan Leigh Anderson, and Beishui Liao. "An approach for combining ethical principles with public opinion to guide public policy." *Artificial Intelligence* 287 (2020): 103349.

Reinforcement Learning:

Cranefield, Stephen, and Ashish Dhiman. "Identifying Norms from Observation Using MCMC Sampling." In Proc. of the 30th International Joint Conference on Artificial Intelligence. International Joint Conferences on Artificial Intelligence. 2021.

Data Mining:

Savarimuthu, Bastin Tony Roy, Stephen Cranefield, Maryam A. Purvis, and Martin K. Purvis. "Obligation norm identification in agent societies." *Journal of Artificial Societies and Social Simulation* 13, no. 4 (2010): 3.

Savarimuthu, Bastin Tony Roy, Stephen Cranefield, Maryam A. Purvis, and Martin K. Purvis. "Identifying prohibition norms in agent societies." *Artificial intelligence and law* 21, no. 1 (2013): 1-46.





• What if the observed emotion is incomplete?

 $joy_i(a,?)$

- Consider the consequences of action a.
- Guessing is defeasible and might conflict with further observation.
- Argumentation to resolve conflicts.

а

Agent i cares about his broken cup because he is distressed after I dropped the cup.

Agent i cares about his coffee because he is distressed after I dropped the cup.



Agent i does not care about his coffee because he is not distressed after I grab his coffee.



- A quantitative approach $B_i(\operatorname{Val}_1(\varphi) \land \ldots \land \operatorname{Val}_t(\varphi) \land \operatorname{Val}_{t+1}(\varphi)) \leftrightarrow B_i O \varphi$
- How to handle exceptions? Argumentation



I have observed one case of $\neg Val_i(\varphi)$, so I believe $\neg O\varphi$.

Thanks for your attention!