What Do You Care About: Inferring Values from Emotions

Extended Abstract

Jieting Luo
University of Bern
Bern, Switzerland
jieting.luo@unibe.ch

Mehdi Dastani
Utrecht University
Utrecht, Netherlands
M.M.Dastani@uu.nl

Thomas Studer
University of Bern
Bern, Switzerland
thomas.studer@unibe.ch

Beishui Liao
Zhejiang University
Hangzhou, China
baiseliao@zju.edu.cn

ABSTRACT

Observers can glean information from others' emotional expressions through the act of drawing inferences from another individual's emotional expressions. It is important for socially aware artificial systems to be capable of doing that as it can facilitate social interaction among agents, and is particularly important in human-robot interaction for supporting a more personalized treatment of users. In this short paper, we propose a methodology for developing a formal model that allows agents to infer another agent's values from her emotion expressions.

KEYWORDS

Emotions; Values; Modal Logic; Argumentation; Multi-agent Systems

1 INTRODUCTION

As it has been recognized that emotions play an important and functional role in driving individuals' behavior, some efforts have been made in artificial intelligence to provide formal models of emotions [1–3, 11, 12]. Another important dimension of emotions is the social aspect, which concerns emotion expression of one agent influences the behavior of another agent with whom it interacts. In particular, observers can glean information from others' emotional expressions through a process that is called inferential processes by van Kleef in [14]. Such information includes the inner state of the agent who expresses emotions, her social intentions, and her orientation toward other agents. For instance, if an agent believes that another agent is joyful for buying ice cream, the observer agent will infer that the expresser agent likes ice cream. It is important to implement the mechanism of inferential processes in socially aware artificial systems as it can facilitate social interaction among

agents, and is particularly important in human-robot interaction for supporting a more personalized treatment of users.

However, inferring an agent's values from her emotion expressions is challenging [8]. Current emotion-recognition approaches based on machine learning are able to distinguish different types of emotions from facial expressions, but are not capable of inferring the agent's inner state that causes the emotion [5][10]. One intuitive approach to this problem is to allow the observer agent to connect an observed emotion with her background knowledge about the event that has triggered the emotion. For instance, if an agent believes that another agent is joyful after buying ice cream, and the observer agent has the belief that buying ice cream will result in having ice cream, she can infer that the observed agent desires to have ice cream. In this paper, we propose a methodology for developing a formal model that allow agents to infer another agent's values from her emotions expressions. It is essential in understanding and modeling the inferential process of emotions in social settings and paves the way for the design and development of software agents that are able to establish affective interaction with human users.

2 LOGICAL FRAMEWORK

As we aim at modeling interaction scenarios, we consider a multi-agent setting consisting of agents, actions and states. A transition relation represents the dynamics of the system and it is labeled with an agent's action. An epistemic accessibility relation for each agent i is used to represent agents' belief. According to appraisal theory in psychology, emotions result from people's evaluation of their perceived events. Different agents may have different emotional reactions to the same event, depending on their standards they use for evaluation. For example, kids gets joyful for having dessert while diabetes patients gets distressed for having dessert. Thus, we can use personal values as agents' evaluation standard, which are similar to goals and desires. In order to simplify our methodology, we assume that a value v is a literal in a finite set L and agent i's procession of value v, expressed as $\text{Val}_i(v)$, is a state property. Having this framework allows agents to reason about actions, beliefs of the environment and another agent's procession of values.

3 MODELING EMOTIONS

In order to model the emotion theory that agents share for their affective interaction, we adopt the OCC psychological model of

emotions where emotions are structured based on *focus of attention* [9]. Values are used as an agent's internal evaluation standard as in [6]. With this psychological model of emotions, we can define an agent i's emotion with respect to an action a and a value v as a state property. For example,

$$\text{joy}_i(a, v) \overset{def}{=} B_i(v \wedge \langle -a \rangle \neg v) \wedge \text{Val}_i(v),$$

which means that agent i is joyful with respect to action a and value v if and only if agent i believes that it is the case that v holds and v did not hold before action a was performed and agent i has value v. An emotion contains important information about expressers' belief of the environment and their values based on which they evaluate states and actions. In reality, however, it is possible that the observer agent observes an emotion while the reason behind is not clear. For example, an agent observes that her user is distressed for dropping a cup of coffee, but the agent is not sure what the user cares about in this incident. This is called an incomplete emotion in this paper. For example,

$$\text{ijoy}_i(a) \overset{def}{=} \bigvee_{v \in L} \left(B_i(v \wedge \langle -a \rangle \neg v) \wedge \text{Val}_i(v) \right),$$

which means that agent i is joyful with respect to action a if only if there exists a literal v in L such that agent i believes it is true and it was false before action a was performed and v is agent i's value.

4 INFERENTIAL PROCESS

If an agent observes and believes the emotion expression with complete information from another agent, the observer agent can infer the expresser's belief of the environment and her value based on which she evaluates states and actions. For instance, if an agent believes that another agent is distressed about getting rice, the former agent can infer that the latter agent does not like eating rice; if an agent believes that another agent is hopeful for getting her paper accepted, the former agent can infer that paper acceptance is important for the latter agent. Such a process is called *inferential processes* in [15]. What if the observer agent observes an incomplete emotion? The action that have triggered the emotion is given, but the consequence of the action that the emotion is directed at is unknown. In such a case, the observer can infer this internal information through connecting the observed incomplete emotion with her background knowledge about the consequence of action. For example, a robot observes that her user gets distressed for dropping a cup of coffee, and from this incident the robot can infer that her user cares about the cup or the coffee inside the cup. An incomplete emotion results in disjunctions, which is a form of ambiguity, but *new* information that is inferred over time can be used to exclude some disjuncts. If the robot continues to observe her user's emotions, one day she might find out that her user is not distressed for dropping a cup without coffee inside. From this the robot can infer that the user does not care about the cup, and he can conclude based on previous information that the user cares about the coffee. As we can see, the observer agent infers and collects useful information from her observation history, and thus she needs an approach to handle all this information for disambiguating another agent's values.

Formal argumentation is a nonmonotonic formalism for representing and reasoning about conflicts based on the construction and the evaluation of interacting arguments [4]. If two arguments have opposite conclusions, then they attack each other. A maximal consistent set of arguments can be found using Dung's well-known argumentation semantics. In particular, structured argumentation frameworks such as ASPIC$^+$ and ABA allows us to build an argumentation theory on the meta level over classical logic [7][13]. We thus propose to use formal argumentation to resolve conflicts between arguments as a way for the observer agent to disambiguate another agent's values. An observer agent constructs an ordinary argument, supporting a value v belongs to an expresser agent, when the observer agent believes an incomplete emotion with respect to an action from the expresser agent and that the action usually brings about a consequence v. An observer agent can also construct a blocking argument, opposing a value v belongs to an expresser agent, when the observer agent believes that the expresser agent does not express a specific incomplete emotion with respect to an action but believes that the action usually brings about a consequence v. The observer agent can then specify the attack relation based on what an argument supports and opposes. Following the interpretation of the arguments, two ordinary arguments attacks each other if and only if they support different values for an incomplete emotion that is observed in the same state, because only one emotion is observed in a state and we do not allow an agent to express an emotion for two different values given that a value is assumed to be a literal. One blocking argument attacks one ordinary argument if and only if the value that the blocking argument opposes is supported by the ordinary argument. We now can construct a Dung-style abstract argumentation framework AFV with respect to an observation history and an agent's values. Because an AFV is defined with respect to an observation history, once the observer agent performs one more observation, she can construct more arguments regarding the expresser agent's values and more attacks and add them to the current AFV. Under the grounded semantic, the observer agent can calculate a set of arguments that are critically acceptable together, which indicate a set of values that are believed to belong to the expresser agent and a set of values that are believed not to belong to the expresser agent. But it should be emphasized that the result only holds with respect to the observation history, because it might change as the observer agent constructs more arguments due to her further observations and updates the argumentation framework accordingly.

5 CONCLUSIONS

This paper proposes a methodology for modeling the mechanism of influential processes that allows agents to infer another agent's values through observing her emotional expressions. It can be used for building agents' Theory of Mind, which refers to the capacity to understand other individuals by ascribing mental states to them. Moreover, it can be extended for the inference of social and ethical values among a society, which has close connection with social norms, and thus future work can be done on the identification of social norms from emotion expressions based on our model. Another direction can be to enrich and advance our model for inferring another agent's value *preferences* apart from her procession of values from her emotion expressions.

REFERENCES

[1] Carole Adam, Andreas Herzig, and Dominique Longin. 2009. A logical formalization of the OCC theory of emotions. *Synthese* 168, 2 (2009), 201–248.

[2] Mehdi Dastani and Emiliano Lorini. 2012. A logic of emotions: from appraisal to coping.. In *AAMAS*. 1133–1140.

[3] Mehdi Dastani, Emiliano Lorini, John-Jules Meyer, and Alexander Pankov. 2017. Other-condemning anger= blaming accountable agents for unattainable desires. In *International Conference on Principles and Practice of Multi-Agent Systems*. Springer, 15–33.

[4] Phan Minh Dung. 1995. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. *Artificial intelligence* 77, 2 (1995), 321–357.

[5] Andrius Dzedzickis, Artūras Kaklauskas, and Vytautas Bucinskas. 2020. Human emotion recognition: Review of sensors and methods. *Sensors* 20, 3 (2020), 592.

[6] Jieting Luo and Mehdi Dastani. 2022. Modeling Affective Reaction in Multi-agent Systems. In *Proceedings of the 21st International Conference on Autonomous Agents and Multiagent Systems*. 1681–1683.

[7] Sanjay Modgil and Henry Prakken. 2013. A general account of argumentation with preferences. *Artificial Intelligence* 195 (2013), 361–397.

[8] Desmond C Ong, Jamil Zaki, and Noah D Goodman. 2019. Computational models of emotion inference in theory of mind: A review and roadmap. *Topics in cognitive science* 11, 2 (2019), 338–357.

[9] Andrew Ortony, Gerald L Clore, and Allan Collins. 1990. *The cognitive structure of emotions*. Cambridge university press.

[10] Lin Shu, Jinyan Xie, Mingyue Yang, Ziyi Li, Zhenqi Li, Dan Liao, Xiangmin Xu, and Xinyi Yang. 2018. A review of emotion recognition using physiological signals. *Sensors* 18, 7 (2018), 2074.

[11] Bas R Steunebrink, Mehdi Dastani, and John-Jules Ch Meyer. 2009. A formal model of emotion-based action tendency for intelligent agents. In *Progress in Artificial Intelligence: 14th Portuguese Conference on Artificial Intelligence, EPIA 2009, Aveiro, Portugal, October 12-15, 2009. Proceedings 14*. Springer, 174–186.

[12] Bas R Steunebrink, Mehdi Dastani, and John-Jules Ch Meyer. 2012. A formal model of emotion triggers: an approach for BDI agents. *Synthese* 185, 1 (2012), 83–129.

[13] Francesca Toni. 2014. A tutorial on assumption-based argumentation. *Argument & Computation* 5, 1 (2014), 89–117.

[14] Gerben A Van Kleef. 2009. How emotions regulate social life: The emotions as social information (EASI) model. *Current directions in psychological science* 18, 3 (2009), 184–188.

[15] Gerben A Van Kleef. 2016. *The interpersonal dynamics of emotion*. Cambridge University Press.