Modeling Affective Reaction in Multi-agent Systems

Extended Abstract

Jieting Luo
Zhejiang University
Hangzhou, China
luojieting@zju.edu.cn

Mehdi Dastani
Utrecht University
Utrecht, the Netherland
M.M.Dastani@uu.nl

ABSTRACT

Affective reaction is a mechanism that is triggered by a perceived emotion and results in the observer's emotion, subsequently steering the behavior of the observer. It is crucial to understand this mechanism as it constitutes a mediator between the emotional expressions of one agent and the (behavioral) responses of another thus facilitating social interaction among agents. This paper reports on a formal specification of affective reaction based on its psychological analysis using values and value systems as agents' appraisal standards.

KEYWORDS

Emotion; Multi-agent system; Logic; Value; Human-robot interaction

ACM Reference Format:

Jieting Luo and Mehdi Dastani. 2022. Modeling Affective Reaction in Multi-agent Systems: Extended Abstract. In *Proc. of the 21st International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2022), Online, May 9–13, 2022*, IFAAMAS, 3 pages.

1 INTRODUCTION

In our daily life, we may get emotional after perceiving the emotion of another person. For example, when walking in the street, we may feel fearful when seeing an angry pedestrian around us, making us to get away immediately. According to the book written by Gerben A. van Kleef, the mechanism underlying the social effects of the pedestrian's anger on us is called affective reaction [7]. It is triggered by a perceived emotion and results in the observer's emotion, subsequently steering the behavior of the observer. Traditionally, emotions were seen as irrational behavior that disrupts rationality. It has been replaced by the pro-emotion consensus that claims that emotion is complementary to rational thinking and behavior, and in support of these [1][4]. We use these insights to design artificial systems because it can help not only to get artefacts with a more human-like behavior thus having deeper and more meaningful human-machine relationship, but also to direct agents' attention to what is relevant, important and significant in order to ensure effective behavior [5] [2][3]. We expect to implement the mechanism of affective reaction in socially aware artificial systems as it can facilitate social interaction among agents, which is particularly important in human-robot interaction for supporting a more personalized treatment of users. Imagine a social context where a robot is serving a patient to drink medicine, but the patient drops his cup onto the ground by accident so his cup gets broken and the medicine splashes out. The patient feels distressed about the broken cup because he cannot use the cup any more. For a fully rational robot without emotion, he will keep filling the cup with medicine for the patient as he has a goal to do that, ignoring the emotion of the patient. For a functional robot that can deal with emotions, he will feel joyful once he observes the distress of the patient and knows a way to cheer up the patient, and this complimentary feeling may lead to bringing a new cup for the patient before filling the cup with medicine again, if he values the patient's emotion. In order to implement such a mechanism into computational models, it is important to have a logical model that allows us to understand its components and their logical relations. In this paper, we propose a model of multi-agent systems using values and value systems as agents' appraisal standards. Different types of emotions such as joy, distress, hope and fear are defined with respect to an action and a value. We then provide a formal definition of affective reaction which takes a perceived emotion of another agent as an input and results in an emotion of the observer agent, subsequently affecting the behavior of the observer agent to cope with his emotion.

Proc. of the 21st International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2022), P. Faliszewski, V. Mascardi, C. Pelachaud, M.E. Taylor (eds.), May 9–13, 2022, Online.

2 LOGICAL FRAMEWORK

This section will introduce the model we use and values and value systems that agents use for their appraisal process. The model we use is an epistemic model consisting of multiple agents Agt, states S, actions Act, a function d that returns the non-empty set of actions in state s, a serial, transitive and euclidean binary relation over states representing the mental state of each agent $\{\mathcal{B}(i)\}_{i \in Agt}$, and a valuation function π. According to appraisal theory in psychology, emotions result from people's evaluation of their perceived information. Different people may have different emotional reactions to the same event, depending on their standards they use for evaluation. For example, someone gets joyful for breaking a cup because he can have a new one; someone gets distressed for for breaking a cup because he cannot use it anymore. In this paper, we use value systems as agents' evaluation standards. A value can be seen as an abstract standard according to which agents have their preferences over states. For instance, if we have a value denoting *equality*, we prefer the states where equal sharing or equal rewarding hold. In this paper, we assume that agents can always compare any two values. Thus, we define a value system as $V = (\mathrm{Val}, \prec)$, where Val is a set of values and $\prec$ is a strict total ordering over Val. In other words, every element in the set of values is comparable to each other and none of them is equivalent to each other.

3 MODELING EMOTIONS

In this section, we will define four types of emotions using a value system as an agent's internal appraisal standards. We adopt the

OCC psychological model of emotions where emotions are structured based on *focus of attention* [6]. At the most general level, all emotions are valenced reactions, each of which is either positive or negative. At the next level, the OCC model specified that valenced reactions can be directed at either consequences of events, actions of agents, or aspects of objects. For this paper, we will model emotions that are directed at consequences of events, which are consequences of physical actions. The same as how we model a value, an emotion is interpreted as a state property, which is represented as a formula. We interpret the positive and the negative aspects of emotions in terms of value promotion and demotion caused by the performance of an action: we say that value v is promoted by action a in state s if and only if v does not hold previously and hold after action a is performed, and we say that value v is demoted by action a in state s if and only if v holds previously and does not hold after action a is performed. As an agent does not always believe that his or her value gets demoted or promoted, so we should refer to an agent's belief instead of objective facts when modeling an agent's emotion.

According to the OCC psychological model of emotions, with respect to valenced reactions to consequences of events, a distinction on emotions is made based on whether the consequence in question is *actual* or *prospective*. The term *actual* is used to refer to both *current* and *certain* events, while the term *prospective* is used to refer to both *future* and *uncertain* events. Many formalizations appear to use OCC's notion of actuality and prospect in only one of these senses. We use certain actuality and uncertain prospect when formalizing emotions as how it is treated in [2] for this paper. Joy and distress are two emotions that are directed to certain consequences of an event, so if an agent is joyful/distressed about the consequence of an event, then he should be certain about the positive/negative consequence of an event. On the contrary, hope and fear are two emotions that directed to uncertain consequences of an event, so if an agent hopes/fears that an event will lead to the desirable/undesirable result, then it is not certain about that. We say that agent i is joyful with respect to value v and action a, if and only if agent i believes that value v does not hold and he believes that action a will bring about v. Dually, we say that agent i is distressed with respect to value v and action a, if and only if agent i believes that value v holds but he also believes that it does not hold after action a is performed.

$$\text{joy}_i(a, v) \overset{def}{=} B_i \neg v \wedge B_i[a]v$$

$$\text{distress}_i(a, v) \overset{def}{=} B_i v \wedge B_i[a]\neg v$$

We say that agent i is hopeful with respect to value v and action a, if and only if agent i believes that value v does not hold and he is uncertain about the effect of action a in terms of v; dually, we say that agent i is fearful with respect to value v and action a, if and only if agent i he believes that value v does not hold and he is uncertain about the effect of action a in terms of v.

$$\text{hope}_i(a, v) \overset{def}{=} B_i \neg v \wedge \text{uncertain}_i(a, v)$$

$$\text{fear}_i(a, v) \overset{def}{=} B_i v \wedge \text{uncertain}_i(a, v)$$

where $\text{uncertain}_i(a, v) = \neg B_i[a]v \wedge \neg B_i[a]\neg v$. We use $e_i(a, v)$ to denote agent i' emotion in any arbitrary type of the above with respect to value v and an action a.

4 MODELING AFFECTIVE REACTION

In a social context, whenever agents perceive the emotional expressions of others around them, their emotions may be influenced by those expressions, which may subsequently inform their behavior. For example, one agent's display of distress may evoke the feeling of distress in an observer, which may result in a desire to provide help to the distressed agent, and expressions of anger may inspire anger and/or fear in an observer, which may motivate fight and avoidance respectively. Such a mechanism where the emotion of an agent is triggered by the emotion of another agent is called *affective reaction* by psychologist Gerben A. van Kleef [7], and it constitutes a mediating mechanism between the emotional expressions of one agent and the (behavioral) responses of another. Affective reaction is the observer's emotion that responds to the expresser's emotion. Thus, the observer perceives the expresser's emotion, which is an important element in the generation of the observer's emotion. In this paper, we do not consider how the observer perceives the expresser's emotion but simply assume that in a given state the observer believes the expresser's emotion. The belief of the expresser's emotion and the belief of the prospective consequence of a given action result in affective reaction. Given two agents i and j, an action b and a state s, the truth of formula $AR_{i,j}(e_j(a, v), b)$ that agent i has affective reaction to agent j's emotion $e_j(a, v)$ with respect to action b is evaluated as follows:

$$AR_{i,j}(e_j(a, v), b) \overset{def}{=} B_i e_j(a, v) \wedge \\ (B_i[b]\neg e_j(a, v) \vee \text{uncertain}_i(b, e_j(a, v))),$$

where $\text{uncertain}_i(b, e_j(a, v)) = \neg B_i[b]e_j(a, v) \wedge \neg B_i[b]\neg e_j(a, v)$ and $e_j(a, v) \in V_i$ or $\neg e_j(a, v) \in V_i$. In words, agent i has affective reaction to agent j if and only if agent i has a value regarding agent j's emotion and he believes that he can or might change it by performing action b. The definition implies that agent i becomes emotional after perceiving agent j's emotion and the reason behind, which corresponds to the notion of affective reaction in the psychological book [7]. Moreover, agent i's belief about agent j's emotion works as the precondition, and agent i's emotion works as the postcondition for this mechanism. In other words, affective reaction can be simply understood as a mechanism with an agent's emotion as input and another agent's emotion as output.

5 CONCLUSION

Affective reaction is a mechanism that is triggered by a perceived emotion and results in observers' emotion. Its social effects and potential to influence the observer's behavior make a logical specification essential for modeling human-like social interaction and designing socially aware artificial systems. In this paper, we have reported the sketches of a logic-based framework that allows to model affective reaction in the context of multi-agent systems. Values and value systems were used as agents' appraisal standards. We defined four types of emotions with respect to an action and a value, based on which a formal definition of affective reaction was provided. In the future, we intend to extend our framework to model more types of emotions that are defined with respect to the history of actions rather than the future consequences of actions and bring the framework closer to programming.

REFERENCES

[1] A.R. Damasio. 1994. *Descartes' Error*. HarperCollins. https://books.google.com.hk/books?id=MocSdi6LXCkC

[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] Nico H Frijda. 2004. Emotions and action. In *Feelings and emotions: The Amsterdam symposium*. 158–173.

[5] Koen V Hindriks and John-Jules Ch Meyer. 2019. Artificial Intelligence in Health Care and Medicine: A Personalized Approach. *Acta scientific medical sciences* 3, 10 (2019), 71–78.

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

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