Hope and Fear in Reinforcement Learning Agents: Extended Abstract[1]

Thomas M. Moerland Joost Broekens Catholijn M. Jonker

Interactive Intelligence, Delft University of Technology

Abstract

We study models of emotion generation in reinforcement learning agents, focussing on anticipatory emotions. Taking inspiration from the psychological Belief-Desire Theory of Emotions (BDTE), our work specifies models of hope and fear based on best and worst forward traces. Results illustrate the plausibility of these signals, for example in the game *Pacman*. Our models enable learning agents to elicit hope and fear, and moreover, explain what anticipated event caused the emotion.

1 Introduction

This paper studies models of anticipatory emotions in reinforcement learning (RL) agents. Computational emotion models are usually derived from the agent's decision making architecture, of which RL is an important subclass. Studying emotions in RL-based agents is useful for different research fields. For machine learning (ML) researchers, emotion models may improve learning efficiency, e.g. by influencing action selection. From a human-robot interaction (HRI) perspective, emotions may communicate agent state (i.e. transparency) or create empathy and enhance user investment. This can help robots and agents transition to domestic environments in the forthcoming years.

To this end we first need plausible models of emotion generation in RL agents (i.e. Markov Decision Process (MDP)-based agents), which is the topic of this paper. Previous work on emotion elicitation in RL agents focussed on model-free learning [1]. However, important emotions like hope and fear are anticipatory, i.e. they require explicit forward simulation. Moreover, forward simulation also allows the agent to explain which anticipated event caused the emotion, i.e. to aid transparency. This work introduces the first anticipatory models of hope and fear in a RL agent. In particular, we show how hope and fear can be efficiently estimated from the best and worst forward traces. Our results show the plausibility of these signals, for example in the game *Pacman*.

2 Method

Models are grounded in the psychological Belief-Desire Theory of Emotion (BDTE) [2]. According to BDTE, hope and fear originate when the *belief* $b(s) \in [0, 1]$ about a state is smaller than 1, while the *desirability* $d(s) \in \mathbb{R}$ is larger or smaller than 0, respectively. We estimate the belief from the path probability towards it, and a positive or negative desirability from the positive or negative temporal difference (TD) error, respectively (see full paper for details). We write s, a, r, γ for state, action, reward and discount parameter, $P(\cdot|s, a)$ for the state transition function, $\pi(s, a)$ for the policy, and $V(s)$ for the state value. Moreover, let a trajectory of depth d from the current node s_0 be given by $g_d = \{s_0 a_0 s_1 a_1 ... s_{d-1} a_{d-1} s_d\}$. The fear in the current node s_0 is then estimated as:

[1]The full paper has been published in: Moerland TM, Broekens J, Jonker CM. Fear and Hope Emerge from Anticipation in Model-Based Reinforcement Learning. In: *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI)* 2016, pages 848-854.

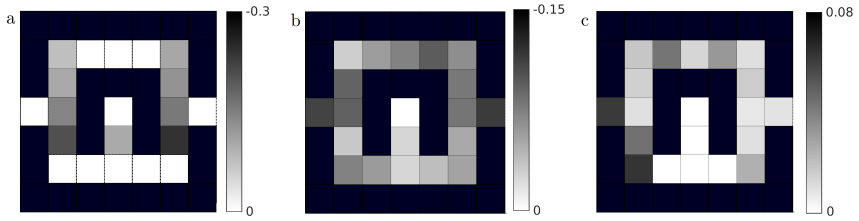

Figure 1: Pacman. a) Fear per location for a ghost *below* and ϵ-greedy(0.10) policy. b) Fear per location for *no* ghost and ϵ-greedy(0.10) policy. c) Hope per location for *no* ghost and softmax(τ=10) policy.

$$F(s_0) = \min_{s'} \left[b(s'|s_0) \times d(s'|s_0) \right]$$
$$= \min_{s'} \sum_{d, g_d | s_d = s'} \left[\prod_{t=0}^{(d-1)} \pi(s_t, a_t) P(s_{t+1}|s_t, a_t) \cdot \left(\left[\sum_{t=0}^{(d-1)} \gamma^t r(s_t, a_t, s_{t+1}) \right] + \gamma^d V(s_d) - V(s_0) \right) \right]^- \quad (1)$$

where $[\cdot]^-$ denotes the negative part. We effectively identify the successor node with highest product of path probability and TD, summing over all possible paths towards it. The model for hope is similar, replacing the $\min$ for a $\max$ and the negative part for the positive. Traces are identified by integrating a Monte Carlo Tree Search procedure (UCT) into the model-based RL architecture.

3 Results

We evaluate our models in *Pacman* (Figure 1). Pacman starts from the center-top, needs to reach the center ($r = +1$), but is chased by a ghost that starts from the center ($r = -1$ for a capture). Pacman can see whether a ghost exists in each cardinal direction, but not the distance to it. We train until Pacman has estimated a stable policy and world model, and then evaluate the plausibility of Pacman's emotions in different scenarios (Figure 1). Figure 1a shows fear when a ghost is seen below. We see how Pacman smoothly learned to be most afraid near the corridor bottoms, as the ghost must then be close. Figure 1b shows fear when Pacman does *not* observe a ghost in any direction. At nearly all locations Pacman fears a capture, but this is most prominent just before the corners (e.g. at the most left and right). Pacman start to feel safer when he approaches the center.

Finally, Figure 1c shows hope when Pacman does not see a ghost. Although we expected Pacman to be hopeful near the center, it turns out to be most hopeful just before the corners, hoping to step around it and still not see the ghost. This implies a jump in the value function, and happens on both sides of the top corridor, at the bottom of the left corridor, and also on the far left (note how Pacman has developed a left-wing tactic). Altogether, these unexpected results illustrate plausibility of the signals, as Pacman identified locations where things might change for better or worse.

4 Conclusion

This paper introduced the first anticipatory models of hope and fear in a RL (i.e. MDP-based) agent. Our emotions emerge from the agent's functionality and a (very sparse) reward signal, without the need for any pre-wired (and ad-hoc) solutions. Future work includes *using* the elicited emotions, for example to benefit learning, e.g. by biasing action selection, or to benefit human-agent/robot interaction.

References

[1] Joost Broekens, Elmer Jacobs, and Catholijn M. Jonker. A reinforcement learning model of joy, distress, hope and fear. *Connection Science*, pages 1–19, 2015.

[2] Rainer Reisenzein. Emotional experience in the computational belief–desire theory of emotion. *Emotion Review*, 1(3):214–222, 2009.