Affect and Learning: a computational analysis

Broekens, D.J.

5

Affect and Modulation

Related and Future Work

In this chapter we discuss other approaches towards computational modeling of affect as well as directions for future work.

5.1 Related Work

The work described in the previous two chapters relates to emotion and motivation based control/action-selection. We explicitly define a role for emotion in biasing behavior-selection as do Avila-Garcia and Cañamero (2004), Cos-Aguilera and others (2005) and Velasquez (1998). The main difference is that in these studies emotion directly influences action selection (or motivation(al states)), while we have studied the indirect effect of emotion-controlled information processing influencing action selection, either by biasing simulation selection (Chapter 4) or by biasing the greediness of action selection (Chapter 3).

A recent variation of this type of research has been presented by Blanchard and Cañamero (2006). In this study, artificial novelty and affect are coupled to exploration behavior of a robot that has to autonomously explore different possible distances to a box. Familiarity (non-novelty) modulated by positive affect is coupled to exploration. The authors argue that their study reproduces behaviors observed in nature. However, their concept of exploration (in contrast to ours) is limited to the single behavioral choice of whether or not the robot should approach the box. This strongly narrows down the meaning of exploration, which is also acknowledged by the authors. Our approach thus contributes to this research by systematically investigating how affect can be used to modulate (mental) exploration in a broader sense.

Two fairly different approaches—different from ours and different from each other—towards studying the relation between affect and adaptive behavior are the work by Lahnstein (2005) and the work by Salichs and Malfaz (2006). Lahnstein shows how the emotive episode (i.e., the short term onset and decay of an emotion) can result from anticipation of reward in the first phase of approaching a reinforced object, while in the second phase the emotive episode is taken over by an evaluation of the actual reward received from that object. There is no space here to do justice to this approach that is important to the process of emotion elicitation in adaptive agents in the spirit of, e.g., Rolls (2000). However, we do want to point out the main difference between Lahnstein's approach and ours, i.e., we use affect in the "mood" (long term) sense as influence on the broadness of mental exploration, while Lahnstein focuses on the process of elicitation of the

short term emotive episode produced by mental anticipation (and reward evaluation). It would be interesting to integrate Lahnstein's result (i.e., the form and elicitation of an emotive episode) with ours, such that our measure of long-term affect is based upon averages over the positive/negative aspect of Lahnteins short-term emotion.

Salichs and Malfaz (2006) introduce an interesting way in which affect can be embedded into the value function Q of a standard Reinforcement Learning method. They enhance Q-learning so that the reward is based on the happiness/sadness of the agent, where happiness and sadness are derived from the agent's wellbeing. Wellbeing is a function over the extent to which the agent's drives are met. So, the more drives met, the happier the agent. This means that their agent is intrinsically motivated by affect, and strives to "maximize happiness" (Salichs & Malfaz, 2006). They use fear, modeled as a parameter that dynamically modulates to what extent the agent chooses—in a world with a stochastic reward function—a risky but optimal policy versus a conservative policy. Fearless agents emphasize actions that are potentially good, while fearful agents more strongly consider the effect of actions that are potentially bad. Their approach thus differs from ours, but, again, both approaches could be integrated such that wellbeing based on drives provides the reward signal and thus our measure for artificial affect is based upon wellbeing averages.

Strongly related to our approach to affect-modulated exploration is research by McMahon et al. (2006), Morgado and Gaspar (2005) and Gadanho (1999; 2003). We discuss this work in more detail in the rest of this section.

McMahon et al. (2006) show how the discrete choice between exploration and exploitation trials can be controlled by a probability value that is derived from measures inspired by affect. This probability uses two measures: one derived from the accuracy of prediction for the upcoming reward as given by the learning mechanism that learns to predict values for future states; the other derived from the actual rewards received. As a result, the probability to explore is high when rewards are low and errors are made in the value prediction, while exploitation is high when rewards are high and prediction errors are low. In this manner they show that agents learn a grid-world problem faster when using this probability value to control exploration. Several interesting differences between their approach and ours should be noted. First, our artificial affect dynamically modulates the amount of mental exploration that influences action selection, while their probability is used for a discrete choice between whether a trial is an exploration or an exploitation trial. Second, their reward-related measure of affect is based on a scaled value for the current reward, where scaling is based on the

min and max rewards obtained in the environment. This means that this measure is unable to model "boredom" (McMahon et al., 2006). Our measure of affect—also related to (the history of) rewards—addresses this issue and is a useful extension to the work of McMahon and colleagues. When our agent has acted in the same environment for a long time, the long and short term averages will converge to the same value and as such artificial affect will be lower, even though the agent might receive huge rewards. In our first hypothesis, low artificial affect results in higher (mental) exploration. This is "boredom" in exactly the same nature as proposed in (McMahon et al., 2006). Third, we have extended the analysis of the psychological plausibility of reward-related measures for artificial affect, which is an issue of future work in (McMahon et al., 2006).

In her PhD thesis, Gadanho (1999) shows an impressive collection of experiments that investigate the relation between affect and adaptive behavior. Here we will discuss several of them. First and foremost, she shows that affect and emotions can be embedded into adaptive agent architectures in a vast amount of ways. These include internally generated emotions as reinforcement to the agent, emotion as interrupting triggers that initiate alternative behaviors when needed, affect-based learning rates, and affect-based exploration/exploitation tradeoffs. Experimental results also vary from positive to negative (in terms of behavioral and learning efficiency). Here we will focus mainly on affect as meta-parameter setting, i.e., affect-based modulation of learning rate and affect-based control of exploration rate (Section 4.6 in Gadanho, 1999).

The experimental setting used is a grid world with an agent that has a neural network robot controller. Every action has a neural network that learns to predict the value of that action in a certain situation. The networks learn with a learning rate η. An action-selection module selects actions, using a temperate parameter T, based on the predictions of the neural networks. The grid world contains walls (avoid), lights (reinforced) and different starting locations (to vary where the robot starts). The goal of the agent is to optimize reward over time. The agent can have four potential emotions, *fear*, *happiness*, *sadness* and *anger*, based on the agents internal drives, *hunger*, *pain*, *body-temperature*, *restlessness* and *eating*. Emotions are elicited continuously during the behavior of the robot. For the current discussion it is not necessary to go into the details of the emotion elicitation model used.

In the emotion-modulates-learning-rate case, the intensity of the agent's dominant emotion is used as gain (multiplication) factor for the learning rate learning η. This gain is equal to 0 if there is no dominant emotion. Experimental results are difficult to interpret. There does not seem to be a generic learning

benefit. This is obvious, as the agent can only learn when it experiences an emotion. Therefore learning will always be slower. However, the results have to be interpreted differently (Gadanho, 1999). One could say that the agent saves learning resources by only learning if its emotion indicates a significant situation. Indeed, in the long run, the agent does learn appropriate food finding behavior, so it is not hindered by slower learning. However, it might be that different tasks benefit more from affect-based learning rate modulation than the task used by Gadanho. Consider the following. In a task where the environment can suddenly change (unlike the task used by Gadanho), an increase in "fear" (e.g., due to running around and not eating) could trigger the start of an intensive learning period. Once the task has been learned again, fear would drop, and the agent would stop learning. While being "bored" (i.e., neutral emotion) the agent will not learn. This is good, as it therefore cannot unlearn previously learned behavior either. In this scenario, it is clear that affective modulation of the learning rate is also useful for adaptation in general, not just for saving learning resources. This thought experiment should, however, be investigated experimentally.

In the emotion-modulates-exploration-rate case, the intensity of the agent's dominant emotion is used to control the Boltzmann temperature. The best results in terms of learning performance are found when a negative dominant emotion is coupled with an increase in temperature. This means that when the agent is feeling bad, it starts to explore. This is exactly the same relation we have found in our studies, and as such there seems to be some convergence on the negative affect=exploration relation. Further, we have extended the studies by Gadanho in the following way. We have studied in more detail the exact behavior of this relation, including many alternative relations between affect and exploration. We have explicitly grounded the relations to the psychological affect and learning literature. Finally, we have explicitly developed several learning tasks (*candy task*, *switch task*) that enable to better investigate the potential of affective modulation.

Morgado and Gaspar (2005) take a slightly different approach. Theirs is more related to our work on affect-bounded thought (Chapter 4), instead of affect-based regulation of exploration in learning (Chapter 3; Gadanho, 1999; McMahon et al., 2006). They present a theoretical framework that explicitly defines a strong, dynamic relation between emotion and cognition. We focus on one aspect of their framework as this aspect relates strongly to our approach: affective bounding of cognitive effort. Affect is derived from how well the agent is doing, analogous to our approach. However, it is defined slightly different. Affect—*Emotional Disposition* as they call it—is defined as a point in a two dimensional space. The dimensions are *change in achievement potential* (δP, achievement potential is the

degree to which an agent can change the current state of affairs in the direction of the goal) and *change in goal conduciveness* (δF, goal conduciveness is the degree of cooperation of the environment regarding goal achievement of the agent). This means that if an agent has a certain goal, and it is steadily moving towards it, affect will be slightly positive overall, as there are no changes in goal conduciveness but there is a constant positive change in achievement potential. Why? Because the agent gets closer and closer and therefore its potential to change the situation to achieve the goal gets larger and larger (assumed that certainty equates potential, hence being close to a goal means a high level of certainty about being able to get to your goal).

In their framework, behavior is understood as the reduction of the difference between the current situation (referred to as *observation*) and a goal situation (referred to as *motivator*). Both observation and motivator are defined as coordinates in a cognitive space, referred to as *cognitive elements*. The dimensions of this cognitive space are equal to the many different qualities a cognitive element can have. For example, dimensions could include "intensity of sunlight", "outdoor temperature", "sea and waves", "having interesting books around" and "cocktail availability". If the current situation is described by a cognitive element having low values on all of these dimensions, while a motivator cognitive element (a goal) exists that has high values on all of these dimensions, one needs a sun-and-beach vacation. The distance between the observation and motivator points thus indicates how much behavior is needed in order to get to a goal state.

Affect relates to cognition and behavior in the following way. Decrease of distance between the points defining current situation and goal state relates to an increase in achievement potential, and increase in speed with which the agent moves towards the goal state relates to increase in goal conduciveness. If I book my vacation and am on the road, my potential to achieve my goal increases, and there is a sudden increase of the goal-achieving speed (acceleration of my observation cognitive element in the cognitive space with respect to the motivator element). As such, positive affect relates to positive δP and δF, while negative affect relates to negative δP and δF.

Affect influences information processing by regulating attentional effort of the agent in two ways. Affective signals are integrated over time into a dynamic threshold ε. If a cognitive element has an activation value (let's call this its saliency) that is higher than the threshold ε, it is considered in the thought process of the agent. Second, a similar threshold ω is used to control the amount of processing the agent has available before it needs to act. Affect thus controls what

cognitive elements enter working memory, as well as how long these elements can stay there for active contemplation. It should be clear that there is a strong analogy with our approach. With regards to affect-based control of resources, the key differences are:

First, we have defined affect in terms of average reward signal changes, while Morgado and Gaspar (2005) exclusively define affect in terms of goal-orientation. This is rather limited, as it assumes a purely cognitive interpretation of affect elicitation and it needs a representation of a future goal. For their purpose this might be sufficient, but for modeling more down-to-earth effects on learning and adaptation (such as the influence of affect on exploration versus exploitation), defining affect in terms of goals is problematic. Representations of future goals might simply not be available.

Second, we specifically—and more elaborately—relate our measure of affect to psychological studies that relate to the influence of affect on learning and information processing. While it is clear that our measure of affect is simple (and in many aspect simplistic), it is strongly grounded in psychological findings. This cannot be said of affect as per Morgado and Gaspar (2005). For example, it is not clear in their model what it actually means if δP and δF do not have the same sign, nor is it very clear how emotions relate to the four possible quadrants of δP and δF.

Third, our approach focuses on the influence of affect-based control of the amount of processing (i.e., internal simulation) on the *learning effectiveness* of an adaptive agent, while Morgado and Gaspar (2005) focus on *problem solving effectiveness* by coupling their affect-based control mechanism to a standard planner. The planner was embedded into an agent that had to continuously plan routes to changing food location in order to maximize food intake. The main result of their experiment is that affect-based control can indeed make more efficient use of planning resources than non-affect-controlled planners.

Regardless of the differences, their results are promising and complementary to ours. They show that even when starting from a very different theoretical point of view, affect-based control of information processing can be a useful method to help resource-bounded agents adapt.

5.2 Future Work

The maximum total amount of simulation used in the setups in Chapter 4 could be fixed, while affect controls *when* to simulate. Now, experiments can be conducted to completely control for the generic effect of the positive influence of more simulation on learning. Arousal could control simulation by, e.g., controlling the

depth of anticipation (or the forgetting rate of the memory so that arousal influences the adaptation speed of the memory).

Even though affective control of exploration versus exploitation seems promising for adaptive behavior and is compatible with psychological findings, our learning model is specific. This means that our claims are hard to generalize. A good way to further investigate the mechanisms of affective control introduced in this chapter is to use different learning architectures, such as *Soar*, or ACT-R. Using the ACT-R architecture, Belavkin (2004) has shown that affect can be used to control the search through the solution space, which resulted in better problem-solving performance. Belavkin has an information-theoretic approach towards modeling affect that is related to the rule state of the ACT-R agent. A key difference is thus that our artificial affect is based on a comparison of reinforcement signal averages. Further we have explicitly modeled affect according to different theoretical views on the relation between affect and information processing and compared these different views experimentally. The "Salt" model by Botelho and Coelho (1998) relates to Belavkin's approach in the sense that the agent's effort to search for a solution in its memory depends on, among other parameters, the agent's mood valence.

As Soar has recently been extended with RL mechanisms, called *Soar_RL* (Nason & Laird, 2004), it is becoming a good candidate for adaptive behavior research. First, *Soar* is a well-understood architecture. Second, *Soar* allows many forms of planning, enabling a better comparison between affective control of planning versus forward internal simulation. We are currently investigating the affect-based control techniques introduced in Chapter 3 and 4 in *Soar_RL* (Hogewoning et al., 2007).

Affective control should be investigated in other types of learning environments, as different environments have their own set of difficulties and particularities for action selection and learning, and imply different functions and benefits for emotion (Cañamero, 2000). Also, more complex and more realistic tasks should be used to test the affect-based mechanisms proposed in the previous two chapters.

On the biological level, there is considerable evidence of the link between positive affect, adaptive behavior and dopamine (Ashby et al., 1999), as well as dopamine, RL, and adaptive behavior (Dayan & Balleine, 2002; Montague, Hyman & Cohen, 2004; Schultz, Dayan & Montague, 1997). Relating our model to this literature is a direction for future work.