Architecture for Emotional Appraisal of Multi-timescale Nexting

Christian Maksymiw, Johannes Feldmaier, Dominik Meyer, Hao Shen and Klaus Diepold
Department of Electrical and Computer Engineering, Institute for Data Processing, Technische Universität München

***Abstract*— In human-machine interaction it is beneficial if artificial agents can communicate their actual internal state towards a human observer in a believable and easily acceptable way. To this end, we propose to extend the architecture of a learning robotic system by a component to determine and display emotions to reflect the internal state.**

I. Motivation

Human beings as well as other animals seem to use experiences from earlier situations to anticipate what is about to happen next. This process of continuously anticipating the immediate future in a local and personal sense is called *nexting*. Modayil et al. [1] have already shown a technical implementation of nexting on a mobile robot. This robot was able to learn how to simultaneously predict all its raw sensor signals at different timescales in real time. They achieved multi-timescale nexting by using the temporal difference learning algorithm TD(λ) with function approximation. Moreover, they also showed how to determine the performance of the multi-timescale nexting algorithm after completion of the learning process in a quantitative manner. These quantitative measures can be interpreted by experts in the field. However, non-experts are often incapable to do so.

In order to enable also non-experts to assess the performance of the multi-timescale nexting algorithm, or in other words to assess the reliability of the computed predictions in a qualitative way during the learning process (online), we take one aspect about nexting into consideration that has been ignored up to our knowledge: nexting is never an isolated process in human beings. Indeed, the contrast between the outcome and our expectations has a significant influence on our current emotional state. If our expectations are exceeded, we feel highly delighted. In contrast, if the outcome falls short of our expectation, some humans are surprised or interested, whereas others could even become disappointed or annoyed.

II. Proposed Architecture

Due to the link between nexting and emotions in human beings, we make use of a concept based on current appraisal theories of emotion in order to simplify the assessment of the reliability of predictions performed by multi-timescale nexting for non-experts. Analogous to human emotions which emerge from the individual's own evaluation of the immediate, imagined, or remembered situation, we use Scherer's component process model (CPM) [2] to map the output signals of the multi-timescale nexting process onto a two-dimensional pleasure-arousal space. The resultant point in this space can then be mapped onto distinct emotional expressions which convey the current affective state of the agent. Subsequently, this representation can be displayed with a suitable emotion display and easily interpreted by a non-expert.

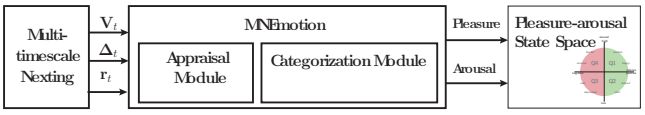

Fig. 1. A simplified version of the general archictecture of MNEmotion and its connections to the preceding multi-timescale nexting process and the following depiction in the pleasure-arousal state space. This space (adapted from [3]) has two possible affective states (core affects) in each quadrant.

Figure 1 illustrates the general architecture which we refer to as MNEmotion. It links to the preceding multi-timescale nexting process as well as to the following representation of the current affective state of the agent in the pleasure-arousal state space.

III. Contribution and Impact

The main goal of the current study is to build an autonomous agent capable of giving direct emotional feedback according to its current internal state. For this purpose we have combined the multi-timescale nexting process with a model of emotion. With this combination we deliver one component of the agent's system of emotion and cognition.

The results show that the agent is able to express online its current affective state during the learning process with the feelings of pleasure and arousal. These feelings are mapped to distinct emotional expressions which can be easily interpreted by non-experts in order to evaluate the prediction process.

There are at least two applications for this kind of affective evaluation of a multi-timescale nexting process. Firstly, the agent can display its current *personal opinion* about the accuracy of the preceding multi-timescale nexting process by using emotions. This is one of the fastest ways for communicating the current learning state of an agent in a dynamic and interactive environment to a human. Secondly, the agent could use the affective representation of the accuracy of predictions for regulating the multi-timescale process itself, i.e. adapting its learning rate parameters.

References

[1] J. Modayil, A. White, and R. S. Sutton, "Multi-timescale nexting in a reinforcement learning robot," in *Adaptive Behavior - Animals, Animats, Software Agents, Robots, Adaptive Systems*. Sage Publications, Inc., 2014, vol. 22, no. 2, pp. 146–160.

[2] K. R. Scherer, "The component process model: Architecture for a comprehensive computational model of emergent emotion," in *Blueprint for Affective Computing: A sourcebook*, K. R. Scherer, T. Bänziger, and E. B. Roesch, Eds. Oxford University Press, 2010, pp. 47–70.

[3] J. A. Russell, "Core affect and the psychological construction of emotion," *Psychological Review*, vol. 110, no. 1, pp. 145–172, 2003.