Artificial Autonomous Agents with Artificial Emotions

Luís Miguel Botelho
Department of Information Sciences and Technologies of ISCTE
Lisbon, Portugal
Phone: 351 1 7935000
luis@iscte.pt

Helder Coelho
Department of Computing Sciences
Faculty of Sciences
University of Lisbon
Lisbon, Portugal
Phone: 351 1 7573141
Helder.Coelho@di.fc.ul.pt

1. INTRODUCTION AND MOTIVATION

This extended abstract summarizes work fully described in [1]. The global research goal of our project (Salt & Pepper) is to define architectures that embody mechanisms that play the same roles, in artificial autonomous agents, as those responsible for the success of human beings. In particular, we focus on the roles played by emotions. Artificial counterparts of natural emotions will be called "artificial emotions".

Natural emotions are the key to the successful use of "pure" problem solving and decision-making skills as measured by intelligence tests like the IQ and the SAT. Emotions intervene at a meta-level of control. They help the organism to regulate itself in order to improve its adaptivness. Artificial emotions will do the same. Among other things, emotion can be used as a basis for attention control, for performance evaluation and regulation, for identification of and recovery from malfunctioning-components, and for adaptive learning processes.

In the current state of our research on artificial emotions, we don't address emotions such as fear, grief and anger. We focus on more basic emotions such as attention shift warnings, performance evaluators, and motivation intensifiers.

In [1], emotion is seen as a sequential, possibly iterative process that comprises appraisal stages, signal generation stages and emotion-responses. This process view of emotion served as a basis for an emotion classification scheme and as a framework for emotion representation.

Emotions are classified along five dimensions: the role/function of emotion, the mediating process by which emotion fulfills its role, the urgency of the repairing process, the source of appraisal, and the type of appraisal that elicits emotion. The main advantage of this classification scheme is to explicitly state which components of the emotion process are at stake in each dimension of classification.

Emotion is described in terms of the signals generated by appraisal processes, and in terms of the responses of the organism to those signals. Emotion-signals are represented by label, valence, intensity and urgency. Emotion-responses are represented by programs in long term memory. There is a one-to-many relation between emotion-signals and emotion-responses, that is, emotion is non-deterministic.

2. APPRAISAL

We distinguish cognitive from affective appraisal, on an architecture-grounded basis. Cognitive appraisals are performed by the cognitive components of the architecture (cognitive engine). Affective appraisals are performed by its affective components (affective engine). Affective and cognitive processing have essentially the same nature. However, the affective engine can be distinguished from the cognitive engine in terms of the type of information processed, the purpose of information-processing, and the typical processing time. It is worth noting that while a cognitive appraisal consumes cognitive resources (e.g., processing time), an affective appraisal doesn't interfere with the current cognitive task of the agent. Our research has been focused essentially on the affective appraisal process.

The purpose of the appraisal process is to evaluate the state of satisfaction of the motives of the organism (e.g., instincts, needs, desires goals, attitudes, values). In order to improve reaction time, the affective appraisal process should avoid comparisons of the current global state with explicit representations of the motives of the organism. Long chains of general production rules with explicit reference to those motives should be progressively replaced by small and specific productions with no explicit reference to motives. This may be achieved through continuous compilation processes, and through a hierarchical organization of the appraisal rules, in which, general rules are used only if no specific rule is available. Affective appraisals are essentially heuristic processes, in the sense

Autonomous Agents '98 Minneapolis MN USA

that they have access only to a very restricted fraction of the information available. The sensors used by the affective engine should extract only a small set of features of the environment. The affective engine should have access only to a small fragment of long term memory (both procedural and declarative): episodic memory and cognitive representations of affects.

3. SALT & PEPPER AND EMOTION

The Salt & Pepper architecture for autonomous agents displays the same kind of regulation capabilities provided by emotion. Salt & Pepper has three major independent blocks that run in parallel with each other: the Cognitive Engine, the Affective Engine and the Interrupt Manager.

In Salt & Pepper, long term memory is an associative network with spreading activation. Emotion-responses are contained in nodes stored and interconnected in long term memory.

Label, valence, urgency and intensity of emotion-signals are generated by the affective engine through the affective appraisal of the agent's global state. When an emotion-signal is produced, the most activated emotion-response matching the signal receives an extra activation that is a function of the intensity of the signal. If the activated emotion-response is activated enough, the Interrupt Manager may decide to load it in working memory where it will be executed.

4. CLASSIFICATION OF EMOTIONS

Emotions have been classified in a large number of ways by different authors. However some of those classification proposals are different just because different things are being classified. One step towards clarification is to precisely identify what components of the emotion process are addressed in each dimension of classification: the appraisal stage, the emotion-signal or the emotion-response. Table 1 summarizes our proposal. Table 2 describes possible responses for the first two dimensions of emotion classification: the role/function of emotions and the mediating process by which emotions fulfill their roles.

Attention shift warnings (col. 1 of Table 2) are useful when an external event occurs that represents an important opportunity or risk for the agent. Performance evaluators (col. 2 of Table 2) have special interest as a means for either "on-spot" performance improvement, or adaptive learning processes.

5. REFERENCES

[1] Botelho, L.M. and Coelho, H. Artificial emotions, http://iscte.iscte.pt/~luis/artigos/artigo11/emo3.ps, Working Paper

Dimension of classification	Examples	Process component
Role/function of emotion	Attention shift warning, performance evaluation, malfunctioning-component warning, motivation intensifier	Emotion-signal
Process by which emotion fulfills its role	Reflexive action, creation of motivators, setting plan selection criteria	Emotion-response
Urgency of the repairing process	Urgent (e.g., need to immediately attend the external environment), not urgent (e.g., need for long-term improvement of default criteria for plan selection)	Emotion-response
Source of appraisal	External environment, internal state, past events, current events	Appraisal stage
Type of appraisal	Affective appraisal, cognitive appraisal	Appraisal stage

Table 1 - Dimensions of emotion classification

Role / Process	Attention shift warnings	Performance evaluators
Reflexive action	(1) Interrupt current processing, initiate new processing.	(4) Trigger autonomic adaptive learning processes that avoid undesirable outcomes or achieve desirable states of affairs
Motivator creation	(2) Interrupt current processing, create delayed attention shift goal, resume current processing	(5) Generate goals whose fulfillment repairs undesirable outcomes due to poor performance, or maintains desirable states of affairs
Plan selection	(3) Adopt plans with periodic monitoring of the external environment	(6) Adopt plans that contain actions that repair undesirable outcomes, or maintain a desirable state of affairs

Table 2 - Role/function of emotion vs. mediating process