

Artificial Emotions for Artificial Systems

Saman Harati Zadeh, Saeed Bagheri Shouraki, Ramin Halavati

Computer Engineering Department, Sharif University of Technology Azadi St., Tehran, Iran

{harati, sbagheri, halavati}@ce.sharif.edu

Abstract

To produce emotional artificial systems in AI domain, usually a subset of human emotional states are imported to the target domain and the major differences between natural and artificial domains are often ignored. In this paper we will discuss about why such an approach is not useful for all possible applications of emotions and we will show how it is necessary and possible to produce artificial emotion systems based on the target systems goals, abilities and needs.

Introduction

The efforts to generate emotions in machines, started by presenting models of human's emotion-arousal mechanism. These models provide a specific relation between states that the subject experiences and the emotions arising in those states (Bower81, Colby81, Pfeifer82, Ortony88, Rollenhagen89, Scherer92, Pfeifer98). A state is defined by a set of features that usually affect the emotion arousal process in human beings. In other words, in most of these models the process of activating specific emotional states is emphasized while usually the functional role of emotions in decision-making process is ignored. In AI domain, such models are the most important sources of inspiration for researchers to promote emotional artificial systems. This approach to emotions in AI has caused the artificial emotional systems be extremely dependent to human-specific emotional states and their arousal mechanisms. In this paper we will discuss how the dependence to the limited set of human emotional states can avoid us from using emotional mechanisms in a vast variety of artificial systems. Also we will propose an alternative approach for modeling artificial emotions and we will show where it can be more useful than the current methods used for modeling emotions.

Human Emotions in Artificial Systems

Most of the current emotion-related research in AI domain is based on the psychological findings about human emotion mechanism and emotional states. Emotions have a long history in psychology domain. Numerous theories and models have been proposed to explain the emotional aspects of human's mind (Freud17,

Davitz69, Frijda96). In recent decades some psychologists have tried to propose models of emotions to be implemented in computers, in order to simulate the emotion mechanism of human. However in these efforts, the emphasis is on the set of emotional states of human and the way they are aroused, and less attention has been paid to the functional role of emotions in decision making process. Possibly because of the ease of implementation, the models of human's emotion arousal mechanism have gained much attention among AI researchers and have been implemented (fully or partially) frequently in artificial systems' framework. On the other hand, the *theories* of emotions discussing the role of emotion system as an element of decision-making are not as accurate and clear as the *models* and have been hardly used in artificial systems. As a result of this trend in AI, the current emotion mechanisms designed for artificial systems, regardless to the goal for which they are designed, resemble a subset of human being's emotional states (Elliot97, Garcia04, Jiang07). This could be helpful if the system's goal is to simulate some features of natural emotional organisms, however it is not the sole target of AI. The emotion mechanisms help us to make better decisions, promote our social behavior, and increase our chance to survive in long term (Damasio94). These facts encourage the AI researchers to apply emotions in machines in order to increase their degree of *intelligence*. The question that may arise is that "Is it necessary to resemble the emotional states of human beings in machines in order to make them more intelligent?" or "Is the human-specific emotional states the most suitable set of emotions for every artificial system?". The natural emotions are the result of long term evolutionary processes with the sole target of increasing human beings survival chance in their environment. Therefore, the emotional mechanisms of human have been adjusted according to his needs, and abilities. Therefore one can expect that for a system working with totally different concepts, having different needs and abilities and trying to achieve different goals, possibly the ideal set of emotions could be not the same as that of human beings. In addition, sometimes the emotions of human can not be easily defined in artificial systems. For example it could be hard to assign the fear or anger emotions to a system that is processing images and removes the noise. These problems have affected the emotion-related research in AI in several ways, some of which are as follows:

- 1-Human's emotional states are modeled based on high level concepts that are not definable in simple systems. Therefore, complex systems with concepts similar to those existing in human's life, should be applied as test beds for artificial emotion mechanisms.
- 2-Usually the core of emotional system, that is responsible for mapping states to emotions, and emotions to behavioral patterns, is defined directly by the designer, and usually the system is not able to learn or adjust these hardcode mechanisms.
- 3-The emotion mechanisms defined in such a way usually play the role of heuristic solutions for specific problems and therefore, are extremely dependent to the knowledge of the designer about the details of the problems that the system may face and their possible solutions.
- 4-Hardcode emotion definition avoids the systems from forming their own emotional identities and personalities. This could be a big flaw especially if we want to model the social roles of emotions.

As we mentioned, the main reason for these problems is that there is not an accurate human-independent definition or functional model of emotions that could be easily applied in AI domain. Having such a definition or model is necessary if we want to go farther than just simulating a small subset of the features of the human beings' emotions. In the next section, we will discuss about the aspects that a definition of emotions possibly should have in order to be applicable in AI domain.

Defining Emotions in Artificial Systems

In the last section, we concluded that the main problem in current models of emotions that are applied in AI domain is that they explain the aspects of *emotional states* instead of the functional role of *emotion system*. This has caused the models to be completely dependent to a limited set of emotional states of human beings, and also has made them inflexible, complicated and domain dependent. To solve this problems it seems necessary to find a definition for emotions that is based on the concepts existing in artificial systems instead of those that are human-specific. Can we expect that a concept like emotion that still does not have an accurate definition for its natural form, be definable according to the simple concepts of artificial systems? Possibly the same question has been asked for the concept of *intelligence* in the beginning of artificial intelligence's life. Although still we do not have a commonly accepted definition for intelligence, for its natural or artificial form, a general understanding for the intelligence in artificial systems has been achieved that is no longer dependent to the human-specific aspects of intelligence. The AI community has focused on those features of natural intelligent systems that make them distinct from unintelligent ones.

The same approach could be applied for emotions; that is the main question should be "what is the difference between an emotional and non-emotional system?" instead of "what are the emotional states of human and how they are aroused or how they affect the behavior?" Not many theories have been presented that discuss about the differences between being and not being emotional. However it seems that the model presented by Minsky [Minsky06] provides a suitable explanation. Minsky envisages the mind as a cloud of resources: a diverse set of processes and other elements that can affect the way in which the mental activities are performed. At each moment, the situation that the person is experiencing causes a series of activation and deactivation of resources; resources that are the most suitable for being used in current situation are activated and form the current configuration of mind. Such a configuration is called a mental state. With this definition, an emotional state could be interpreted as some specific mental state. Therefore one can assume that an emotion-enabled system should be reconfigurable system according to the situations that it experiences. However being able to support more than one possible configuration is not enough for a system to be called emotion-enabled. Although this condition relaxes the dependability of the artificial emotional systems to human-specific emotional states, still it does not emphasis on ability of the system to form its own set of emotional states. As we discussed earlier, hardcode emotion systems are inflexible mechanisms that play the role of domain dependent heuristics for solving specific problems while the emotion mechanisms in natural organisms are the result of long term evolutionary processes and the knowledge formed by millions of generations. Therefore the natural emotion mechanisms are general and adaptable tools that help the species to survive in a wide range of possible situations that are not known in advance for any single organism. Although learning a suitable emotion mechanism could be a hard problem, it seems that emotions learnt by a system could be much more useful than hardcode mechanisms defined directly by the designer. In addition for an artificial system, emotions could be definable based on concepts that are completely different from those that define our emotional structures and mechanisms. Therefore, it could be impossible for a human designer to predict the suitable emotional patterns for the system. Therefore it is necessary for an emotion-enabled system to be able to form its own emotion mechanism. To do so it has to assign appropriate working configurations to different situations, and build a mechanism to activate these configurations when corresponding situations arise. In the next section we will present an abstract model for an emotion-enabled system and also we will show how it presents the required features and how it could be applied in artificial systems.

An Abstract Model for Emotional Systems

Assume the emotion-enabled intelligent system "M". It should include the following components:

- 1-A set of emotional states $E = \{E_0 \dots E_n\}$
- 2-A set of resources $R = \{R_0 \dots R_m\}$
- 3-An emotion activation mechanism ("EA") that determines the next arousal value of each emotional state (E_i) according to their current arousal values and the current state (S_i).
- 4-An emotional reconfiguration mechanism ("ER") that defines the active resources (R_i) for each emotional state.
- 5-An emotion learning ("EL") mechanism that adjusts the "EA" and "ER" mechanisms.

Such a system could be shown as (E, R, EA, ER, EL). Consider that the exact behavior of the system can be determined by the active resources in each moment.

This model presents most of the aspects that one can expect from an emotion-enabled system. First of all, it covers the hardcode emotional mechanisms that are commonly used in AI: In a hardcode emotion-enabled system "EL" mechanism is inactive. If a system is build based on such a model it will be able to present the following capabilities:

- 1-Generating different behavioral patterns while the system is experiencing different emotional states.
- 2-Supporting multiple policies in evaluating the situations, goals and concerns
- 3-Improving the performance of the system by defining the emotional states as most useful working configurations.
- 4-Emotions are learnt, therefore different experiments can lead to different definitions of emotions. In other words, each system will have an emotional personality that is solely dependent to its own past experiences, and could be completely different from those of other systems.

It is important to note that the model does not dictate any policy to emotion activation and emotional reconfiguration mechanisms and it is the responsibility of emotion learning mechanism to define these policies. However a measure is required for evaluating the performance of the system for different possible emotion definitions. This measure could be defined based on the specific needs or goals of the system. In natural emotion mechanisms, this measure has been the ability of the organism and specie to survive in long term. However for artificial systems, it could be any other measure like the short term or long term reward, or the possibility of system failure. To show how this abstract model can be applied, in the next section we will present a sample system designed using it.

A Sample Emotional System

Imagine an agent that lives in an environment, performs some task and receives reward as the result of its behavior. The agent learns from its experiences and gradually increases its performance. Such a pattern could be found in many areas of AI domain, like reinforcement learning and case based reasoning.

To add emotionality to such a system we define the elements mentioned in last section:

- 1-Emotional states: The exact definition of emotional states are learnt by the system, however the number of possible emotional states could be defined in advance. For example we can assume that the system can support two emotional states: E_0 and E_1 .
- 2-Resources: The system's resources could be defined in different levels of abstraction. In this sample we use the following arrangement:
 - a. The knowledge base used for learning and decision making. In this sample we assume that the system can have one knowledge base for each emotional state, so that the knowledge learnt in each emotional state is stored separately. (R_0 and R_1).
 - b. The Decision Making and Learning mechanism (R_2). We assume that this mechanism does not change in different emotional states.
- 3-Emotion activation mechanism: This mechanism is responsible for computing the activation level of each emotional state in each moment. In this sample we make two assumptions:
 - a. The system can have pure emotional states (that is the most activated emotional state will be the emotional state of the system).

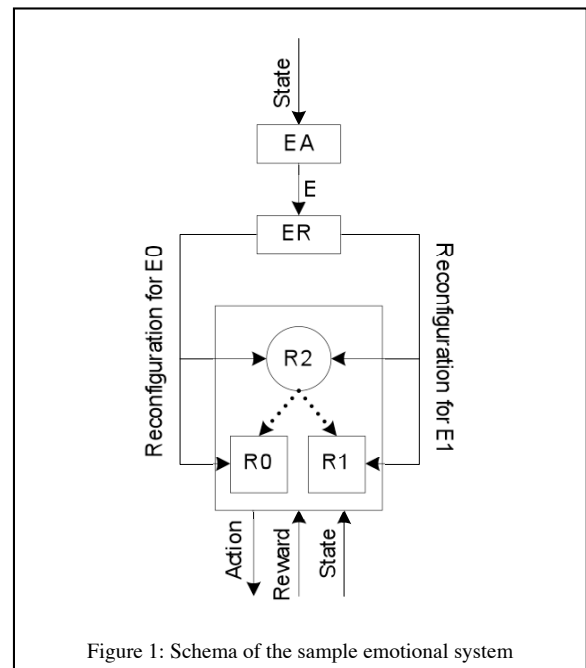


Figure 1: Schema of the sample emotional system

b. The next emotional state is defined independently from the current one.

4-Emotion reconfiguration mechanism: This mechanism defines which resources are active when the system is in some emotional state. Since the resources that could be assigned to different emotional states are of the same type and completely similar, there is no need to learning for this mechanism. We assume that E0 will activate resources R0 and R2 while E1 will activates resources R1 and R2. As it is clear, the reasoning mechanism will be active for all emotional states, but the knowledge base that it uses, will change.

5-Finally the emotion learning mechanism should be defined. As we mentioned, in this sample, the only mechanism that should be learnt is the emotion activation mechanism. In other words the system should learn when to activate each emotional state. To do so, some performance measure should be defined for the system. A simple measure could be the momentary reward that the system receives after each action. The learning mechanism will strengthen the relationship between the state S_i and the emotion E_j if activating E_j in S_i leads to a positive reward, while receiving a penalty weakens this relationship. Such a learning mechanism will lead to emotional states that help the agent to increase the average reward it receives in long term.

The presented sample (figure 1) shows how the abstract model of emotional systems could be implemented. As it is clear, the model provides a straightforward approach for building emotion-enabled systems. Our experiments have shown that such an approach to emotions not only could help us to generate emotional systems, but also can be a tool for increasing different aspects of the intelligence.

Summary and Conclusion

In this paper we discussed about some flaws of the current approach to modeling emotions in AI. We proposed that human-independent definitions for emotions can help us to generate emotions according to the artificial systems' requirements. Also we presented such a model and show how it could be applied in a sample system. The authors believe that to use the full functionality of the emotion mechanisms, including the role of emotions in promoting the performance, forming the personality or providing a communication mechanism, applying human-independent models of emotions is inevitable. We hope that the presented model could be a first step toward achieving better understanding about the concept of emotions in artificial systems.

References

- Avila-Garcia, O., Canamero, L., (2004), Using hormonal feedback to modulate action selection in a competitive scenario. In S. Schaal, A. J. Ijspeert, A. Billard, S. Vijayakumar, J. Hallam and J. A. Meyer (Eds.), *From Animals to animats 8: Proceedings of the eighth international conference on simulation of adaptive behavior (SAB)*, (pp 243–254), Cambridge, MA: MIT Press.
- Bower, G.H. (1981), *Mood and Memory*, American Psychologist, 36.
- Colby, K.M. (1981). Modeling a paranoid mind. In: *The Behavioral and Brain Sciences*, 4 (4), S. 515-560.
- Damasio, A., (1994), *Descartes' Error: Emotion, Reason, and the Human Brain*, New York :Avon Books.
- Davitz, J.L. (1969), *The language of emotions*. New York: International Universities Press.
- Elliott, C. (1997), I picked up catapla and other stories: A multimodal approach to expressivity for "emotionally intelligent" agents. In: *Proceedings of the First International Conference on Autonomous Agents*.
- Freud, S. (1917), *Introductory lectures on psychoanalysis*, *The complete psychological works of Sigmund Freud*. London: Hogarth Press.
- Frijda, N.H. (1996), *Passions: Emotions and socially consequential behavior*, In R.D. Kavanaugh, B. Zimmerbeing & S. Fair (eds), *Emotion: Interdisciplinary Perspectives*, NJ: Lawrence Erlbaum. and pathology of the emotions. *Psychological Review*, 71.
- Jiang, H. (2007), *From Rational to Emotional Agents*, PhD thesis, Department of computer science, University of south Carolina.
- Minsky, M.,(2006), *The Emotion Machine: Commonsense Thinking, Artificial Intelligence, and the Future of the Human Mind*, Simon & Schuster.
- Ortony, A., Collins, G.L. & Clore, A. (1988), *The cognitive structure of emotions*, NewYork, Cambridge University Press.
- Rollenhagen, C. und Dalkvist, J. (1989). *Cognitive contents in emotion: A content analysis of retrospective reports of emotional situations*. Technical Report, Department of Psychology, University of Stockholm.
- Scherer, K. (1992), On social representations of emotional experience: Stereotypes, prototypes, or archetypes? In M. Von Cranach, W. Doise & G. Mugny(eds), *Social representations and the social bases of knowledge*, Burn:Huber.
- Pfeifer, R. (1982). *Cognition and emotion: an information processing approach*. Carnegie-Mellon University, CIP Working Paper Nb. 436.
- Pfeifer, R. and Verschure, P.F.M.J. (1998). *Complete autonomous agents: a research*.