Emotions in a Cognitive Architecture for Human Robot Interactions

Antonio Chella
Dipartimento di Ingegneria Informatica
Universitá degli Studi di Palermo
Palermo, Italy

Ignazio Infantino
Istituto per il Calcolo ad Alte Prestazioni
Consiglio Nazionale delle Ricerche
Palermo, Italy

Abstract

A robot architecture is proposed in which cognitive models of emotions are modelled in terms of conceptual spaces. The architecture has been implemented in a anthropomorphic robotic hand system. Experimental results are described related to an experimental setup in which the robot system plays Rock Paper Scissor against a human opponent.

Introduction

A robot system that acts in the real world has to deal with rich and unstructured environments and it has to interacts with other robots and people. To appropriately act, a robot must be able to deeply *understand* its perceived environment. Developing such a capability requires both *bottom-up*, data driven processes that associate symbolic knowledge representation structures to the data coming out of a vision system, and *top-down* processes in which high level, symbolic information is in its turn employed to drive and further refine the interpretation of the scene.

We claim that affective computing has a main role in the focusing and reinforcement processes during high level understanding of dynamic scenes. In this line, many cognitive theories of emotions have been proposed in the psychological literature, full of suggestions for robot vision systems (Schachter & Singer 1962), (Leventhal & Scherer 1987), among others. However, the implementations proposed in the robotic literature, though impressive, model emotions at a low level of abstraction (Murphy *et al.* 2002), (Velàsquez 1999), (Cañamero & Fredslund 2001), (Michaud 2002), (Breazeal 2002). Briefly, the emotional system of a robot is typically responsible of the tuning of parameters of the robot reactive behaviors.

Our architecture integrates in a simple and principled way artificial vision, artificial emotions and symbolic knowledge representation by introducing a rich and expressive intermediate kind of representation where affective computing takes place. Such *conceptual* kind of representation is situated between *subconceptual* knowledge (extracted by the artificial vision algorithms) and *linguistic* symbolically organized knowledge. More in details, the implemented emotional system drives the robot attention on the conceptual entities which are more relevant according with the current perceptions and the corrent emotional state. Moreover, the emotional system guides the learning of the *hot components* (Barsalou 1999) of the current emotional state by generating a suitable reinforcement signal.

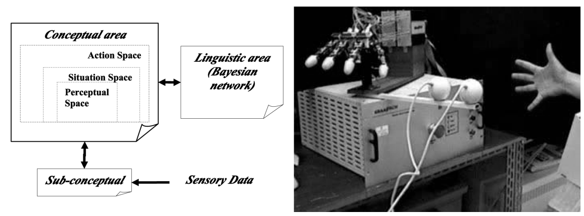

Figure 1: (left) Outline of the cognitive architecture. (right) The robotic hand at work.

The architecture has been implemented and it controls the robotic hand system operating at the Robotics Laboratory of the University of Palermo. The system (see Fig. 1 right) is made of an anthropomorphic robotic hand with four fingers built by GraalTech (Genova, Italy) and a low-cost stereo pair.

The system takes as input a sequence of images corresponding to subsequent phases of the evolution of the scene (the movements of the human hand and their effects on the whole scene), and it generates a suitable action performed by the robotic hand according with the current input, the system inner knowledge and the emotional state. In order to test the system and to have quantitative data on human system interaction, we have analyzed a measurable experimental setup in which the user plays the Rock-Paper-Scissors game against a human opponent, comparing the performances of the whole architecture when the emotional system is activated or not.

The Cognitive Architecture

The proposed robot architecture is organized in three computational areas (Chella, Frixione, & Gaglio 1997), (Chella, Frixione, & Gaglio 2000). Fig. 1 (left) shows the relations among them.

The three areas are concurrent computational components working together on different commitments. There is no privileged direction in the flow of information among them.

In the implementations of functional capabilities of emotions in the proposed cognitive architecture, we hypothesized that emotions are responsible for the expectations of future rewards and punishments of the agent. The main role of emotions is to generate contexts in which the architecture resources are suitably allocated according to the maximization of the agent expectation of rewards (Rolls 1999), (Balkenius & Moren 2001).

The Subconceptual Area

The *subconceptual area* processes the visual data coming out from the video camera. It provides the 3D geometric parameters describing the observed dynamic scene. In this respect, the operations of the subconceptual area are typical computer vision processes, such as edge detection, object tracking and segmentation, resolution of ambiguities and occlusions and so on. In particular, in the implemented setup, this area performs a 3D reconstruction of the acquired hand able to individuate the orientation and reciprocal position with other body parts (Infantino *et al.* 2003).

In this area also resides the emotional system of the architecture that associates the emotional response with each perceived entity, according with the current emotional state. The operation of the emotional system is therefore similar to an associative memory with an inner state. In the current implementation we adopted a set of *gated dipole* neural networks (Grossberg 1988) that learn to associate emotional responses with perceptual information coming from the vision system.

The Conceptual Area

The conceptual area is based on the theory of conceptual spaces (Gärdenfors 2000). and it has the role to link the subconceptual and the symbolic areas. A conceptual space CS is a metric space whose dimensions are related with the quantities processed by the robot sensors. Dimensions do not depend on specific linguistic descriptions; examples could be color, pitch, volume, spatial co-ordinates.

A *knoxel* k is a point in CS and it represents the epistemologically primitive element at the considered level of analysis. The distance $d(k_1, k_2)$ between k_1 and k_2, according to the given metric, has the meaning of similarity between the entities represented by k_1 and k_2. Particular knoxels $\{kc_1, kc_2, \ldots\}$ have the meaning of prototypes of concepts $c_1, c_2, \ldots$: i.e., a particular knoxel may correspond to the prototype of a typical phalanx, another knoxel corresponds to the prototype of some typical object and so on. The concept knoxel are the basis points for the Voronoi tesselation of CS, according with the given distance.

The *prominence* of a concept is modelled by means of a circle in CS centered on the corresponding prototype kc_i with radius r_i. If the concept c_1 is more prominent than c_2 in the current situation, then $r_1 >> r_2$ holds: i.e., the prominence circle of kc_1 is greater than the prominence circle of kc_2, according with the *generalized* Voronoi tessellation.

In order to represent moving parts, objects and actions, we adopts three related conceptual spaces: the *perception space* PS, the *situation space* SS and the *action space* AS.

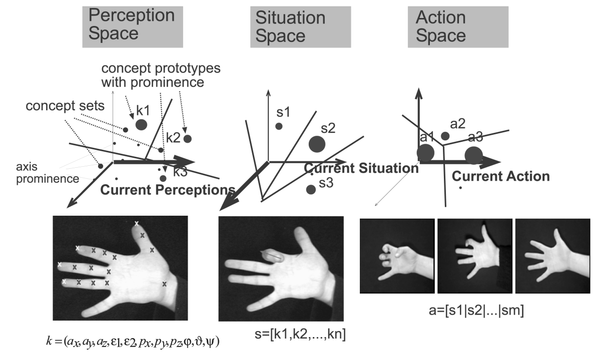

Figure 2: The perception space, the situation space and the action space and the relationship among them.

The perception space PS represents the entities coming from the subconceptual area. Starting from the acquired variation in time of the geometric primitives (e.g., joint angle values) describing the human hand, we adopt a PS in which each point represents a whole simple motion of the components of the hand. In the line of the approach described in (Marr 1982), we consider a simple motion as a motion interval between two subsequent generic discontinuities in the motion parameters. We adopt as PS a function space in which each axis corresponds to a harmonic component according with the Fourier Transform Analysis. In this case, a knoxel is a point in the conceptual space, and it corresponds to the simple motion of a part of the hand (e.g, a phalanx).

Let us now consider the whole hand when the index is opening. It may be represented in PS by the set of the knoxels corresponding to the simple motions of its components. In the case of the hand, some knoxel corresponds to the moving phalanxes of the index, while the other knoxels correspond to the phalanxes of quiet fingers. Each point in the situation space corresponds to a collection of points in PS. We call *situation* this kind of scene and we represent it as a knoxel in *situation space* SS. A knoxel in SS therefore corresponds to the set of knoxels in PS related with the considered situation.

In a situation, the motions of all of the components in the scene occur simultaneously, i.e. they correspond to a single configuration of knoxels in PS. To consider a composition of several motions arranged according to a temporal sequence (e.g., when the hand moves one finger, then another finger and so on), we introduce the notion of *action*. An action is represented as a "scattering" from one situation to another, i.e., from the old situation knoxel in SS to the new situation knoxel in SS, and so on. We represent the action as a knoxel in *action space* AS. A knoxel in AS corresponds to the set of knoxels in SS related with the considered action.

The Linguistic Area

The role of the linguistic area in the proposed architecture is twofold: on the one side, it summarizes the situations and actions represented in the conceptual spaces previously described by suitable linguistic terms anchored to the structures in the conceptual spaces. On the other side, it stores the symbolic, “a priori” knowledge of the system.

In the current implementation of the architecture, the linguistic area is based on the situation calculus formalism (Reiter 2001), that allow us to generate a linguistic representation of dynamic scenes and robot actions. The main motivation for choosing the situation calculus lies in the fact that it is one of the simplest, more powerful and best known logic formalisms for the representation of knowledge about actions and change.

In particular, terms related with the parts of the perceived objects, as the phalanxes, are anchored with entities in PS; terms related with situations, i.e., whole objects with their motion state as the moving hand acquiring a posture are anchored with entities in SS, and terms related with actions, i.e., sequences of hand postures are anchored with entities in AS. In the current version, we adopt the probabilistic situation calculus, that allows us to represent the human hand situations and actions in terms of probabilistic finite state automata.

Robotic Emotions and Conceptual Spaces

In the implementations of functional capabilities of emotions, we hypothesized that emotions are responsible for the expectations of future rewards and punishments of an emotional agent. A main role of emotions is to generate contexts in which the agent resources are suitably allocated according to the maximization of the agent expectation of rewards (Rolls 1999), (Balkenius & Moren 2001).

With reference to the described architecture, emotions are generated by an emotional system in the subconceptual area and they affect the conceptual area. In particular, the role of the emotions is to dynamically change the prominence of the prototypical knoxels in the conceptual spaces, i.e., with knoxels describing parts in PS, with knoxels describing situations in SS and with knoxels describing actions in AS. The changes of prominence are related with emotional responses according to the current emotional state of the system. In this way, the prominence of the concepts related with the current emotional context is bigger than the prominence of the other concepts, in order to let the agent to focus on the entities really important in the current emotional context, i.e., the entities that may generate future rewards and/or avoid future punishments.

The emotional response depends on the current robot curse of actions. For example, in the pleasure state, the robot operations are all normal and all the concepts have their default prominence. If, for some reason, the robot is in the arousal state, the prominence of potentially dangerous entities grows up; in this case a moving object towards the robot deserve more attention that other moving entities.

Emotions may be more or less related with specific concepts: when the robot is in the arousal state, its emotional system generates a response for all of the perceived moving objects. Instead, in the fear state, the emotional system generate a response only for the possibly dangerous objects. In the former case, the emotional system affects the prominence circles of all of the concept knoxels, while in the latter case, the emotional system only affects the prominence circles of the concept knoxels corresponding to some moving objects.

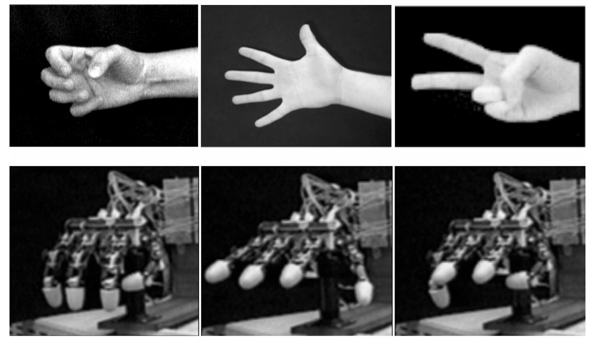

Figure 3: The postures of the human and robotic hand during the RPS game.

The emotional system employed in the current implementation is based on a set of *gated dipole* neural network (Grossberg 1988) that learn to associate emotional responses with perceptual information coming from the vision system. Each gated dipole is related with a concept knoxel. The output of the emotional system is the emotional response associated with concept knoxels. This response in turns influences the prominence circle of the concept knoxel itself. In this way, a concept knoxel with its prominence circle now describes not only the corresponding entity (part, situation or action), but also the related emotional response. The vision system of the robot thus generates the perceptive part of the knoxel and the robot emotional system generates the corresponding emotional response, represented in the corresponding CS by means of the prominence circle.

Experimental Setup

We adopted an experimental setup that allowed us to measure the degree of learning of the system during human-robot interactions. In this setup, human user plays the Rock, Paper, Scissors game against the robotic hand. We have chosen the RPS game because it is simple, fast, involving hand dexterity and strategy between two players. Moreover, RPS game is based on standard hand signs.

On the surface, RPS appears to be a game of chance. Whether because of associations with the symbols or the hand positions that represent them, players perceive the three throws to have distinct characteristics. These vary from player to player, but generally fall into some common patterns. The World RPS Society web site[1] associates different behaviors to the three possible opening throw:

> rock: use of rock as an opening move is seen by many players to be a sign of aggression; paper: it is actually the most challenging of the basic opening moves since it requires the manual displacement of the most digits. It is therefore generally viewed as the least obvious of opening throws; scissors: opening with a pair of scissors assumes that you are playing against an opponent who has tight control over their aggressive tendencies.

[1] `http://www.worldrps.com/`

We may distinguish two main player emotional profiles: the aggressive profile which tend to use rock, and the controlled profile which tends to use scissor. The user of paper is a sort of intermediate between the two profiles. When the system recognize a profile, e.g., the aggressive profile, the prominence of concept knoxels related with rock grows up: i.e., the knoxels related with the rock positions of the phalanxes in PS, the knoxel related with the rock position of the hand in SS and the actions in AS with a prevalence of rock moves. In this case, the system is more able to anticipate rock moves than the other moves.

Another curse of emotion is related with the game: if the system responses are correct, the system is in the pleasure state, and, as before described, the prominence of concept knoxels is related only with the emotional profile of the player. When the system is starting to lose the game, it enters in a concerned emotional state, and the prominence of the current profile start to grow down. When the system is continuing to lose the game, then it enters in a frustrated emotional state and the prominence of concept knoxels change to have a transition between the current profile and the other profile. In both cases, the role of emotions is to suitably allocate system resources in order to maximize expected rewards, i.e., in order to reach the pleasure state.

The choice of the move is based on expectations generated on the basis of the structural information maintained in the linguistic area and described in terms of the probabilistic situation calculus. Essentially, the situation calculus describe the probabilistic state automata implementing the rules of the game. As previously stated, the terms of situation calculus are anchored with entities in the three conceptual spaces, and the current prominence of the concept knoxel, which is related with the current emotional state, is taken into account not only in the recognition of the opponent move, but also in the decision making of the robot move, as it influences the probabilities of the probabilistic automata.

The system has played 500 matches against human user which uses a defined complex strategy based on *gambit* composition. A gambit is a series of three throws used with strategic intent. Strategic intent in this case, means that the three throws are selected beforehand as part of a planned sequence. There are only twenty-seven possible gambits, but they can also be combined to form longer, complex combination moves.

The strategy followed by human player is represented by the union of the gambit (PSR) and (RPR), with the random choose to repeat the same sign or change gambit after a stalemate or the conclusion of a set. A single game uses the best of three of three format (max 3 sets, ended when a player wins 2 throws). In the first phase of the challenge (match #1-#50), the system plays at random, obtained a success rate near to 33% (stalemate is counted as fail).

After approximately 250 matches the system has completely learned the behavior of the human player and has obtained a success rate near to 70%. The various experiments done have highlighted a profile of learning process characterized by a random initial phase that lasts about 75 matches depending from player strategy, a second phase with constant converging learning rate, and a final phase in which the system does not improve its skill.

Results obtained with a similar architecture without the emotional system showed a similar learning profile but a success rate near 61%. The proposed emotion system allows the architecture for a substantial improvement in success rate, mainly because the emotional system make the whole architecture more able to adapt its strategy on line.

References

Balkenius, C., and Moren, R. 2001. Emotional learning: A computational model of the amygdala. *Cybernetics and Systems* 32:611–636.

Barsalou, L. 1999. Perceptual symbol systems. *Behavioral and Brain Sciences* 22:577–660.

Breazeal, C. 2002. *Designing Sociabe Robot*. Cambridge, MA: MIT Press.

Cañamero, L., and Fredslund, J. 2001. I show you how I like you – can you read it in my face? *IEEE Trans. Systems Man and Cybernetics: Part A* 31(5):454–459.

Chella, A.; Frixione, M.; and Gaglio, S. 1997. A cognitive architecture for artificial vision. *Artif. Intell.* 89:73–111.

Chella, A.; Frixione, M.; and Gaglio, S. 2000. Understanding dynamic scenes. *Artif. Intell.* 123:89–132.

Gärdenfors, P. 2000. *Conceptual Spaces*. Cambridge, MA: MIT Press, Bradford Books.

Grossberg, S. 1988. *Neural Networks and Natural Intelligence*. Cambridge, MA: MIT Press.

Infantino, I.; Chella, A.; Džindo, H.; and Macaluso, I. 2003. Visual control of a robotic hand. In *Proc. of IROS'03*.

Leventhal, H., and Scherer, K. 1987. The relationship of emotion to cognition: A functional approach to a semantic controversy. *Cognition and Emotion* 1:3–28.

Marr, D. 1982. *Vision*. New York: W.H. Freeman and Co.

Michaud, F. 2002. EMIB - computational architecture based on emotion and motivation for intentional selection and configuration of behaviour-producing modules. *Cognitive Science Quarterly* 2(3-4).

Murphy, R.; Lisetti, C.; Tardif, R.; Irish, L.; and Gage, A. 2002. Emotion-based control of cooperating heterogeneous mobile robots. *IEEE Trans. on Robotics and Automation* 18:744–757.

Reiter, R. 2001. *Knowledge in Action: Logical Foundations for Describing and Implementing Dynamical Systems*. Cambridge, MA: MIT Press, Bradford Books.

Rolls, E. 1999. *The Brain and Emotion*. Oxford, UK: Oxford University Press.

Schachter, S., and Singer, J. 1962. Cognitive, social and physiological determinants of emotional state. *Psychological Review* 69:379–399.

Velàsquez, J. 1999. An emotion-based approach to robotics. In *Proc. of 1999 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 235–240. Los Alamitos, CA: IEEE Press.