Emotional Human-Computer Interface:

Are emotional inferences forward?

M. Iza[1]

[1] University of Málaga, Málaga, Spain

iza@uma.es

Abstract. The cognitive architecture approach claims for massively parallel data structures and processes. However, none of these models fully address the integration of emotion generation and its effects in the context of cognitive processes. This work tries to unify several models of computational emotions with work done in cognitive architectures. In particular, considering that emotional inferences are forwarded.

Keywords: cognitive architecture, automatic processing; emotional responsivity; computational emotion.

1 Introduction

The cognitive architecture approach claims for massively parallel data structures and processes. Implementing these structures and processes finishes in reduced performance along some dimensions. We need a convergence among cognitive architectures from two points of view: (i) computational aspect, necessity to perform complex tasks in demanding contexts; (ii) implementational aspect, reflecting the exponentially increasing knowledge figure about the nature of brain processes.

Recent cognitive approaches assume the theoretical framework of embodied and situated cognition [1-13]. Within each module there are different kinds of representations and processes. These modules go from perception and action to language understanding and high-level reasoning. The goal is to postulate different cognitive architectures that can explain the interactions between modules. For example, hybrid architectures assume that low level processes are performed by subsymbolic methods and high level processes are performed by AI symbolic methods (e.g., [14-16]).

These architectures mainly refer to the problem of representation. Memory representations are necessary for the activation of cognitive processes [17]. Therefore, the issue is activating representations on long term memory, and bringing them to working memory in order to operate cognitive processes (e.g., Baddeley's model of working memory).

The problem, at this point, is how to deal with emotion within this kind of cognitive architectures. Specifically, if it can be assumed that cognitive processes are previous to emotional processes, or vice versa. In this paper we will discuss this problem considering emotion as a key issue in order to control the extent of cognitive inferences

For instance, [18] present an architecture where three sets of processes can interact: processes responsible for fast context-sensitive behaviors (an autonomous mind), processes responsible for cognitive control (an algorithmic mind) and processes responsible for deliberative processing and rational behavior (a reflective mind). By reasoning on counterfactual situations, the system tries to link emotional semantic and cognition with neuromodulations. These ones, proposed as physiological components, act like an attentional focus on salient emotional aspects of environments.

A possible way for evaluating the provided advancements of the different architectures, is that one of focusing on classes of problems that are easily manageable for humans but very hard to solve for machines. For example, these involve aspects concerning commonsense reasoning about space, action, change and language categorization [19]; selective attention; integration of multi-modal perception; the interaction between cognition and emotion [20,21]; learning from few examples [22]; robust integration of mechanisms involving planning, acting, monitoring and goal reasoning [3,23].

Within this framework, a main hypothesis is that emotional mechanisms play a critical role in structuring the high-level thought processes of cognitive systems. Some models of these mechanisms can be usefully integrated in artificial cognitive systems architectures, which constitute a significant step towards cognitive systems that reason and behave, externally and internally, in accordance with emotional requirements.

However, emotional concepts in these theories are generally not defined formally and it is difficult to describe in systematic detail how processes work. In this sense, structures and processes cannot be explicitly implemented. Some attempts have been incorporated into larger computational systems that try to model how emotion affects human mental processes and behavior [24-27].

As we will see, some tutoring systems have explored this potential to inform user models. Likewise, dialogue systems, mixed-initiative planning systems, or systems that learn from observation could also benefit from such an approach. That is, considering emotion as interaction can be relevant in order to explain the dynamic role it plays in action and cognition.

In this work, we will provide some psychological insights into the emotional grounding of conceptualization and language use. In particular, the role of human-computer interface, in order to develop novel approaches to grounding of robotic conceptualization and language use (more precisely, verbal labelling of objects and actions), based on the insights gained under richer computational and robotic models. We will discuss this problem considering emotion as a key issue in order to control the extent of cognitive inferences.

2 Cognitive architectures

In order to characterize a computational model of emotion, we have to take into account different interdisciplinary uses to which computational models can be put, such as improving human-computer interaction or enhancing general models of intelligence.
Starting from some integrated computational models have tried to incorporate a variety of cognitive functions [10], more recent cognitive systems in AI focus on the role of emotion in order to address control choices by driving cognitive resources on problems of adaptive significance for the agent. For example, human computer interaction attempts to recognize user's emotion including physiological indicators and facial and vocal expressions. Similarly, how we can use of emotion or emotional displays in avatars that interact with the user, for instance, to increase student motivation in a tutoring system.
In this sense, computational models take different frameworks in research and applications. On one hand, psychological models emphasize on fidelity with respect to human emotion processes. On the other hand, AI models evaluate how the modelling of emotion impacts reasoning processes or improves the fitness between agent and its environment. That is, the model improves and makes more effective the human-computer interaction.
Several models have been proposed and developed. However, some fundamental differences arise from their underlying emotional constructs. For instance, as we will see below, some discussions on if emotion precedes or follows cognition disappears if one adopts a dynamic system perspective. Here, we will discuss two main approaches.
On the one hand, some models focus on appraisal as the core process to be modelled. In this sense, emotion is not completely elaborated. Mechanisms for deriving appraisal variables, via if-then rules, model specific emotion label. Here, we can distinguish between a specific emotion instance and a more general affective state. For example, [25] proposed EMA in order to generate specific predictions about how human subjects will afford with emotional situations. An agent that tries to operate in real time, multi-agent environments, would need these appraisal processes. Such as for human computer interaction, these techniques create an interactive agent that deals with emotion.
On the other hand, dimensional theories argue that emotion is not discrete entities. Rather, it is a continuous dimensional space. These theories conceptualize emotion as a cognitive label attributed to a perceived body state, mood or core affect [28]. An agent is considered in an affective state at a given moment and the space of possible states within broad, continuous dimensions.
Although there is a relationship between both approaches, appraisal dimension is a relational construct that characterizes the relationship between some specific event (or

object) and subject's emotion (belief, desire or intention). Even more, several appraisal variables can be active at the same time. Contrarily, the dimension of affect is a non-relational construct, indicating only the overall state of the subject.
These dimensional theories focus on the structural and temporal dynamics of core affect and often do not deal with affect's antecedent in detail. It is conceived as a non-intentional state; the affect is not about some object. Here, despite of symbolic intentional judgments, many sub-symbolic factors could contribute to a change in main affect.

3 Human-computer interface

There have been lots of approaches trying to detail a core set of bases for achieving a human-computer interface. While some of them are more directed to AI applications, others try to point out the design of a psychologically plausible architecture. In any case, these two approaches share many aspects.
For purpose of modeling emotion generation, we have centered on appraisal theories, which are the dominant basis for that type of computational model. Appraisal theories generally argue that people are constantly evaluating their environment, and that evaluations result in emotions such as fear or anger. Each theory differs in its appraisal variables and the way in which appraisals are generated (simultaneously vs. specific order).
Some fusion techniques are required in order to integrate inputs from different modalities. In this concern, several fusion approaches have been developed. In order to support more wide ranging functional multimodal systems, general processing architectures have been developed. They try to joint together a variety of multimodal patterns and their processing.
A typical feature of multimodal data processing is that multisensory data are processed separately and only combined at the end [29]. But, as has been said previously, several inputs cannot be considered in an independent way and must be combined according to a context dependent model and processed in a joint feature space of sensors, cognition and emotion.
This integration has been performed a several levels. On the one hand, some fusion techniques have been applied at the feature level. For instance, in audio-visual integration, one simply can concatenate the audio and visual feature vectors to obtain a combined vector. To reduce the length of this audio-visual vector, dimensionality reduction techniques are applied. The recognition module (e.g., hidden Markov model) can be trained to classify this mixed vector.
There have been proposed some intermediate fusion techniques. Early fusion fails to model the fluctuations in the relative reliability and the asynchrony problems, for example, between the audio and video streams. Often we have to deal with imperfect data in the inputs. This has been achieved by considering the time-instance vs. time-scale dimension of human non-verbal communicative signals [30]. Here, we need some kind of probabilistic inference to manage previously observed data with the current inputs.

Several probabilistic graphical models have been proposed, such as hierarchical hidden Marcov models and dynamic Bayesian networks.
Finally, it is possible to integrate the body of different information at a higher semantic level. We have to fuse common meaning representations derived from different input modalities (sensorial, cognitive and emotional) into an interpretation framework (e.g., audio-visual speech recognition; [31]).
The latter aspect is a crucial issue in order to integrate sensors, cognition and emotion within an agent. Despite important advances, further work is required to investigate this general problem. We could employ individual recognizers that can be trained by using particular data, but they have to interact with a number of input modes or increasing representations. This research area has to address the fusion of heterogeneous input features and combine them in different kind of contexts.

4 Discussion

Models in language processing have researched how words are interpreted by humans. Many models presume the ability to correctly interpret the beliefs, motives and intentions underlying words. The interest relies also on how emotion motivates certain words or actions, inferences, and communicates information about mental state. As we will see below, some tutoring systems have explored this potential to inform user models. Likewise, dialogue systems, mixed-initiative planning systems, or systems that learn from observation could also benefit from such an approach.
As these experimental data show, activating accessible constructs or attitudes through one set of stimuli can facilitate cognitive processing of other stimuli under certain circumstances, and can interfere with it under other circumstances. Some of the results support and converge on those centered on the constructs of current concern and emotional arousal.
Future research has to take seriously into account this question: how to develop models where emotion interacts with cognitive processing. One example could be the work of [32] where it is combined speech-based emotion recognition with adaptive human-computer modeling. With the robust recognition of emotions from speech signals as their goal, the authors analyze the effectiveness of using a plain emotion recognizer, a speech-emotion recognizer combining speech and emotion recognition, and multiple speech-emotion recognizers at the same time. The semi-stochastic dialogue model employed relates user emotion management to the corresponding dialogue interaction history and allows the device to adapt itself to the context, including altering the stylistic realization of its speech.
Interpreting the mix of audio-visual signals is essential in human communication. Researchers have to take into account the advances in the development of unimodal techniques (e.g., speech and audio processing, computer vision, etc). In traditional human-computer interaction, the user faces a computer and interacts with it via a mouse or a keyboard. In the new applications (e.g., multiple agents, intelligent homes) interactions

are not explicit commands. Some of the methods include gesture, speech [31], eye movements [33], etc.
We can interpret the suggested selection mechanism as an information filter. This information filter only selects the measurement for the required features and passes them to the memory system. Features that do not contribute in solving a given task are discarded. This also requires a dynamical and flexible system architecture that allows for a demand-driven combination of processing modules. We have proposed such architecture for the congruent emotion of word processing. To acquire more complex information, the system needs to combine those procedures in a suitable way within memory representation. Beside this, the system has to decide which properties it has to measure for solving the current task. The resulting representation is demand related, as only the pieces of information to solve the task is acquired. This task driven representation can serve as a foundation for learning new relations between words and emotions and for interpreting current interactions.
[20] addressed the problem of the detection and revealing of the relevant "context" to inform affect detection. They implemented a context-based affect detection component embedded in an improvisational virtual platform. The software allows up to five human characters and one intelligent agent to be engaged in one session to conduct creative improvisation within loose scenarios. Some of these conversations reveal personal subjective opinions or feelings about situations, while others are caused by social interactions and show opinions and emotional responses to other participant characters. In order to detect affect from such contexts, first of all a naïve Bayes classifier is used to categorize these two types of conversations based on linguistic cues. A semantic-based analysis is also used to further derive the discussion themes and identify the target audiences for the social interaction inputs. Then, two statistical approaches have been developed to provide affect detection in the social and personal emotion contexts. The emotional history of each individual character is used in interpreting affect relating to the personal contexts, while the social context affect detection takes account of interpersonal, sentence types, emotions implied by the potential target audiences in their most recent interactions and discussion themes. The new development of context-based affect detection is integrated with the intelligent agent.
In this context, a psychological framework of emotional language processing is needed to describe the steps humans take when they interact with other computer systems or agents [26]. This framework can be used to help evaluate the efficiency and naturalness of a user interface (e.g., design principles, emotional inferences, etc.). So, the key question is to represent, reason, and exploit various models of word processing to more effectively process input, generate output, and manage the dialog and interaction between different agents. The input data (words) should be, cognitive and emotionally, processed in a joint feature space according to a context-dependent model.

References

1. M. Boden, AI. Its nature and future. Oxford University Press, 2016.
2. D. Vernon, Artificial cognitive systems: A primer. MIT Press, 2014.

3. A. Clark, Surfing uncertainty: Prediction, action and the embodied mind. Oxford University Press, 2016.
4. G. Dodig-Crnkovic, and R. Giovagnoli, Representation and reality in humans, other living organisms and intelligent machines. Springer, 2017.
5. A. Lieto, M. Bhatt, A. Oltramari, and D. Vernon, “The role of cognitive architectures in general artificial intelligence,” Cognitive Systems Research 2018, 48, 1-3.
6. N. Essau, and L. Kleinjohann, “Emotional robot competence and its use in robot behavior control,” in Emotional engineering, S. Fukuda, Ed; Springer Verlag, London, 2011; pp. 119-142.
7. U. Kurup, and C. Lebiere, “What can cognitive architectures do for robotics?”, Biologically Inspired Cognitive Architectures, 2012, 2, pp. 88-99.
8. T. Ogi, “Emotional design in the virtual environment,” in Emotional engineering, S. Fukuda, Ed; Springer Verlag, London, 2011; pp. 103-118.
9. T. Watanabe, “Human-entrained embodied interaction and communication technology”, in Emotional engineering, S. Fukuda, Ed; Springer Verlag London, 2011; pp. 161-178.
10. M. Anderson, “Embodied cognition: A field guide”, Artificial Intelligence 2003, 149, pp. 91-130.
11. J.R. Anderson, D. Bothell, M.D. Byrne, S. Dougass, C. Lebiere, and Y. Qin, “An integrated theory of the mind”, Psychological Review 2004, 111(4), 10-36.
12. J.R. Anderson, and C. Lebiere, “The Newell test for a theory of cognition”, Behavioral & Brain Science 2003, 26, pp. 587-637.
13. J.E. Laird, The SOAR cognitive architecture. MIT Press 2012.
14. P. Langley, J.E. Laird, and S. Rogers, “Cognitive architectures: Research issues and challenges”, Cognitive Systems Research 2009, 10, 141-160.
15. R. Sun, Duality of the mind: A bottom-up approach to ward cognition. LEA, Mahwah, NJ., 2002.
16. R. Sun, “The CLARION cognitive architecture: Extending cognitive modeling to social simulation”, Cognition and Multi-Agent Interaction 2006, 79-99.
17. R. Sun, Anatomy of the mind: Exploring psychological mechanisms and processes with the clarion cognitive architecture. Oxford University Press, 2016.
18. O. Larue, P. Poirier, and R. Nkambou, “The emergence of (artificial) emotions from cognitive and neurological processes”, Biologicallly Inspired Cognitive Architectures 2013, 4, pp. 54-68.
19. A. Lieto, D.P. Radicioni, and V. Rho, “Dual PECCS: A cognitive system for conceptual representation and categorization”, Journal of Experimental & Theoretical Artificial Intelligence, 2017, 29(2), pp. 433-452.
20. L. Zhanj, and J. Barnden, “Affect sensing using linguistic, semantic and cognitive cues in multi-threaded improvisational dialogue”, Cognitive Computation 2012, 4(4), pp. 436-459.
21. I. Juvina, O. Larue, and A. Hough, “Modeling valuation and core affect in a cognitive architecture: The impact of valence and arousal on memory and decision-making”, Cognitive Systems Research, 2018, 48, 4-24.
22. D. Vanderelst, and A. Winfoeld, “An architecture for ethical robots inspired by the simulation theory of cognition”, Cognitive Systems Research, 2018, 48, pp. 56-66.
23. D. Choi, and P. Langley, “Evolution of the ICARUS cognitive architecture”, Cognitive Systems Research, 2018, 48, pp. 25-38.
24. R.P. Marinier, J.E. Laird, and R.L. Lewis, “A computational unification of cognitive behavior and emotion”, Cognitive System Research 2009, 10, 48-69.
25. S. Marsella, and J. Gratch, “EMA: A process model of appraisal dynamics”, Journal of Cognitive System Research 2009, 10, 70-90.

26. B. Parkinson, “What holds emotions together? Meaning and response coordination”, Cognitive System Research 2009, 10, 31-47.
27. D. Sander, D. Grandjean, and K.R. Scherer, “A system approach to appraisal mechanisms in emotion”, Neural Networks 2005, 18, 317-352.
28. J.A. Russell, “Core affect and the psychological construction of emotion”, Psychological Review 2003, 110, 145-172.
29. A. Jaimes, and N. Sebe, “Multimodal human-computer interaction: A survey”, Computer Vision and Image Understanding 2007, 108, 116-134.
30. M. Pantic, and L.M. Rothkrantz, “Toward and affect sensitive multimodal human-computer interaction”, in Proceedings of the IEEE 2003, 91 (9), 1370-1390.
31. G. Potamianos, C. Neti, J. Luettin, and I. Mattehews, “Audio-visual automatic speech recognition: an overview”, in E Vatikiotis-Bateson and P Perrier (Eds.), Issues in visual and audio-visual speech processing. MIT Press, 2004.
32. J. Pittermann, A. Pittermann, and W. Minker, Handling emotions in human-computer dialogues. Berlin: Springer Verlag, 2010.
33. K. Grauman, M. Betke, J. Lombardi, J. Gips, and G. Bradski, “Communication via eye blinks and eyebrow raises: video-based human-computer interfaces”, Universal Access in the Information Society 2003, 2(4), 359-373.