Report

on a

preliminary literature study

on the topic of

Affective AI

Janos Tapolczai

April 3, 2014

1 Recent developments

A review of the literature shows that in recent years there has been, not so much the development of new theoretical models, but rather the application of tried and true technology like machine learning [1] to novel problems. The foundational research that is being done tends — one might suspect, due to the hardness of the general AI problem — to be rather taxonomical, concerned with the mere description of emotions, not on the mechanism of their evocation and effect. Good examples of this are Celso M. de Melo et al. [2], who presented simple emotional models used in the iterated prisoner's dilemma, Chen Yong et al. [3] who dealt with the classifications of emotions, and Marc Schröder et al. [4], who described an *Emotion Markup Language*.

Other popular topics in Affective AI are crowd sourcing ([5] and [6]) and emotion recognition in voice and video — these, however, lie outside of the scope of this study and shall not be discussed here any further.

2 State of the art

In terms of theory, progress has been rather modest over last few years. When discussing the state of the art, we must distinguish between the prevailing views on what emotions *are* — their nature and their taxonomy — and the techniques used to model emotional *processes*.

2.1 Nature of emotions

Even a cursory reading among published papers makes it clear that there is no consent on the exact nature and purpose of emotions. In discussions of these questions, the following positions emerge:

1. Utility functions — in the context of reinforcement learning, emotions are taken to mean *reward and punishment*, and are used to train neural networks [7, 8] .

2. Somatic markers — emotions are regarded as *shortcuts* in cognition: the activation of a marker in a situation leads to an immediate response; this mechanism is activated instead of logical reasoning, or after it has failed [9, 10] .

3. Evolutionary models — Philippe Chassy and Fernand Gobet present a model in line with evolutionary biology, in which an older "emotional CPU" modulates a newer "cognitive CPU" [11]. Minsky [12] presents a similar view in *The Emotion Machine*, in which he equates consciousness simply to the "noise" made by various, vying mental components.

2.2 Taxonomy of emotions

When it comes to a taxonomy of emotion, we must distinguish between low- and high-fidelity models: low-fidelity models make no claim to faithfully capture the notion of affect; they use the term "emotion" almost metaphorically, as a synonym for "reward function", "utility", or "achievement potential". Examples of such models are found in machine learning and neural networks, and in Morgado & Gaspar, who employed a two-axis model of *achievement potential* and *achievement flow* in the context of emotion-driven memory formation [13].

High-fidelity models endeavour to give an experientially/neurologically faithful picture of emotions and are therefore much more profoundly informed by (a) empirical research and (b) the underlying cognitive model. From research, various authors (Ortony & Turner, Matsumo & Ekman) have derived the existence of six so-called *basic emotions*: happiness, surprise, sadness, fear, disgust, anger[1] [14, p. 24]. These are said to be both qualitatively different from each other and prior to other, derivative emotions like contentment, shame, or aesthetic appreciation.

When it comes to categorizing emotions, various vector-based solutions are popular:

1. Tomkins partitions emotions into *positive* and *negative* affects [15].

2. Davidson & Irwin and Gable & Harmon-Jones, starting from the idea that TODO, propose a different 2-axis model, with emotions being *approach-/avoidance*-related on one axis and positive/negative on the other [16, 17].

3. Breazeal creates a three-axis model with the axes *arousal*, *valence* and *stance* [18]. In this model, the experience of emotions is preceded by an affective appraisal stage, performed via somatic markers (see Damasio's *somatic marker hypothesis*).

Other models build on nested categories instead of axes: Ortony et al. [19] created the OCC model, which contains 22 emotions and categorizes them by valence, consequences for others and oneself, well-being, attribution, etc. This categorization is furthermore hierarchical: being *distressed* is a special case of being *displeased* where one's well-being is in danger; *remorse*, in turn, is a special case of distress, where the cause of displease is then attributed to oneself. Such an inheritance is stressed in [20].

2.3 Models of emotional processes

Models of differing depth and grounded in various philosophies exist to model emotional processes. Most of these, as said, are at least a decade old, although people like Sabrina Campano et al. [21] and Tibor Bosse et al. [22] have recently produced interesting innovations.

1. Neural networks — neural networks can model emotions in two ways: first, they can simulate ordinary neural processes in the hope of producing something like affect as the end result or, second, emotion can be encoded into them as reward and punishment, as described by Rolls[7] and Hyungil Ahn et al. [8]

2. Layered architectures - Minksy [12] describes a layered architecture with higher-lever mental components being stacked onto evolutionarily older, more primitive ones.

3. Hormonal Emotion Architecture — Sandra Clara Gadanho and John Hallam [23] model the mechanics of emotion using *hormones*, which are emitted whenever an agent experiences a certain emotion. In turn, they are fed back into the system and, combined with external stimuli, decide which emotions an agent feels.

4. Chassy's and Gobet's model — as discussed above, Chassy and Gobet view cognition as a rational, cognitive CPU influenced by an emotional component. Perhaps more of a meta-model, this idea could be used to integrate an emotion simulation program with a logical reasoner.

5. COR — Sabrina Campano et al. [21] describe the *COR (Conservation of Resources)* model, based on the acquisition of desired and the protection of threatened resources.

6. Emotion Contagion Spiral — Tibor Bosse et al. [22] and Gonçalo Pereira et al. [24] develop and build upon the *Emotion Contagion Spiral Model*, which aims to simulate the mechanism by which agents mimic and internalize the emotions of their peers. As Chassy's and Gobet's, this is somewhat of a meta-model, but can be useful in modelling multi-agent scenarios.

7. OCC model — The OCC model was invented by Ortony et al. [19]. It introduced an extensive taxonomy, together for rules for when an agent would experience certain emotions. Dastani et al. [25] extended it with a BDI (belief, desire, intention) model grounded in doxastic logic.

---

[1]Matsumo & Ekman additionally include contempt as a basic emotion.

References

[1] Omar Alzoubi, Md.Sazzad Hussain, Sidney D'Mello, and RafaelA. Calvo. Affective modeling from multichannel physiology: Analysis of day differences. In Sidney D'Mello, Arthur Graesser, Björn Schuller, and Jean-Claude Martin, editors, *Affective Computing and Intelligent Interaction*, volume 6974 of *Lecture Notes in Computer Science*, pages 4–13. Springer Berlin Heidelberg, 2011.

[2] CelsoM. Melo, Peter Carnevale, Dimitrios Antos, and Jonathan Gratch. A computer model of the interpersonal effect of emotion displayed in a social dilemma. In Sidney D'Mello, Arthur Graesser, Björn Schuller, and Jean-Claude Martin, editors, *Affective Computing and Intelligent Interaction*, volume 6974 of *Lecture Notes in Computer Science*, pages 67–76. Springer Berlin Heidelberg, 2011.

[3] Chen Yong and He Tong. Affective computing model based on rough sets. In Jianhua Tao, Tieniu Tan, and RosalindW. Picard, editors, *Affective Computing and Intelligent Interaction*, volume 3784 of *Lecture Notes in Computer Science*, pages 606–613. Springer Berlin Heidelberg, 2005.

[4] Marc Schröder, Paolo Baggia, Felix Burkhardt, Catherine Pelachaud, Christian Peter, and Enrico Zovato. Emotionml - an upcoming standard for representing emotions and related states. In *ACII (1)*, pages 316–325, 2011.

[5] Abe Kazemzadeh, Sungbok Lee, Panayiotis G. Georgiou, and Shrikanth S. Narayanan. Emotion twenty questions: Toward a crowd-sourced theory of emotions. In D'Mello et al. [27], pages 1–10.

[6] Laurel D. Riek, Maria F. O'Connor, and Peter Robinson. Guess what? A game for affective annotation of video using crowd sourcing. In *ACII (1)*, pages 277–285, 2011.

[7] Edmund T. Rolls. A Theory of Emotions, Its Functions, and Its Adaptive Value. *Emotions in Humans and Artifacts*, 1:11–32, 2002.

[8] Hyungil Ahn and Rosalind W. Picard. Affective-cognitive learning and decision making: The role of emotions. April 2006.

[9] Antonio R. Damasio. *Descartes' Error: Emotion, Reason, and the Human Brain*. Gosset/Putnam Press, New York, 1994.

[10] Antoine Bechara and Antonio R. Damasio. The somatic marker hypothesis: A neural theory of economic decision. *Games and Economic Behavior*, 52(2):336–372, 2005.

[11] Philippe Chassy and Fernand Gobet. A model of emotional influence on memory processing. *Symposium on agents that want and like: Motivational and emotional roots of cognition and action. AISB 2005*, 2005.

[12] Marvin Minsky. *The Emotion Machine: Commonsense Thinking, Artificial Intelligence, and the Future of the Human Mind*. Simon and Schuster, New York City, 2006.

[13] Luís Morgado and Graça Gaspar. Adaptation and Decision-Making Driven by Emotional Memories. In Carlos Bento, Amílcar Cardoso, and Gaël Dias, editors, *Progress in Artificial Intelligence*, volume 3808 of *Lecture Notes in Computer Science*, pages 102–114. Springer Berlin Heidelberg, 2005.

[14] Jorge Amory and Patrik Vuilleumier. *The Cambridge Handbook of Human Affective Neuroscience*. Cambridge University Press, 2013.

[15] Tomkins S.S. *Affect imagery consciousness. Volume II: The negative affects*. Springer, New York, 1963.

[16] Davidson R. J. and Irwin W. The functional neuroanatomy of emotion and affective style. 3:11–21, 1999.

[17] Gable P. A. and Harmon-Jones E. Approach-related positive affect reduces breath of attention. 19:476–82, 2008.

[18] Cynthia Breazeal. Emotion and sociable humanoid robots. *Internation Journal of Human-Computer Studies*, 59:119–155, 2003.

[19] Andrew Ortony, Gerald L. Clore, and Allan Collins. *The Cognitive Structure of Emotions*. Cambridge University Press, 1988.

[20] Steuenbrink B. R., Dastani M. M., and J.-J. Ch. Meyer. The OCC model Revisited. In *Proceedings of the 4th Workshop on Emotion and Computing - Current Research and Future Impact*, Paderborn, 2009.

[21] Sabrina Campano, Etienne de Sevin, Vincent Corruble, and Nicolas Sabouret. Simulating affective behaviours : An approach based on the cor theory. In *ACII (1)*, pages 457–466, 2011.

[22] Tibor Bosse, Rob Duell, ZulfiqarA. Memon, Jan Treur, and C.Natalie Wal. A multi-agent model for emotion contagion spirals integrated within a supporting ambient agent model. In Jung-Jin Yang, Makoto Yokoo, Takayuki Ito, Zhi Jin, and Paul Scerri, editors, *Principles of Practice in Multi-Agent Systems*, volume 5925 of *Lecture Notes in Computer Science*, pages 48–67. Springer Berlin Heidelberg, 2009.

[23] Sandra Clara Gadanho and John Hallam. Robot learning driven by emotions. *Adaptive Behaviour*, 9(1):42–64, 2001.

[24] Gonçalo Pereira, Joana Dimas, Rui Prada, Pedro Alexandre Santos, and Ana Paiva. A generic emotional contagion computational model. In *ACII (1)*, pages 256–266, 2011.

[25] Bas R. Steunebrink, Mehdi Dastani, and John-Jules Chr. Meyer. A formal model of emotion triggers: An approach for bdi agents. *Synthese*, 185:83–192, 2011.

[26] Rosalind W. Picard. *Affective Computing*. The MIT Press, 1997.

[27] Sidney K. D'Mello, Arthur C. Graesser, Björn Schuller, and Jean-Claude Martin, editors. *Affective Computing and Intelligent Interaction - Fourth International Conference, ACII 2011, Memphis, TN, USA, October 9-12, 2011, Proceedings, Part II*, volume 6975 of *Lecture Notes in Computer Science*. Springer, 2011.

[28] Jianhua Tao, Tieniu Tan, and Rosalind W. Picard, editors. *Affective Computing and Intelligent Interaction, First International Conference, ACII 2005, Beijing, China, October 22-24, 2005, Proceedings*, volume 3784 of *Lecture Notes in Computer Science*. Springer, 2005.

[29] Jung-Jin Yang, Makoto Yokoo, Takayuki Ito, Zhi Jin, and Paul Scerri, editors. *Principles of Practice in Multi-Agent Systems, 12th International Conference, PRIMA 2009, Nagoya, Japan, December 14-16, 2009. Proceedings*, volume 5925 of *Lecture Notes in Computer Science*. Springer, 2009.

[30] Alexander Feder. Methods of affective artificial intelligence and an application for traffic simulations with emotional drivers. Master's thesis, Vienna University of Technology, 2008.