Eyes for an Agent

Simulation of Dynamically Evolving Preferences for Visual Stimuli by a Neural Network and a Multiple-Trace Memory Model

Marius Raab, Mike Imhof and Claus-Christian Carbon
Department of General Psychology and Methodology, University of Bamberg, Germany

Motivation

Our goal is to simulate the dynamics of aesthetic preferences – the well-observable fact that liking judgements are subject to periodic changes. Carbon (2010) showed such a cyclic pattern for car design, with a striking predomination of curved car bodies up to the 70ies (like in the famous Volkswagen *Beetle*); that were replaced by angled styles in the 80ies (as in the first VW *Golf*); which were succeeded by curved designs in the 90ies (as with the VW *New Beetle*).

In our view, these shifting assessments can only be explained when accounting for *motivation* and *emotion*. Everyday-statements that an old design looks *boring* or a new one *thrilling* indicate the emotional impact of design.

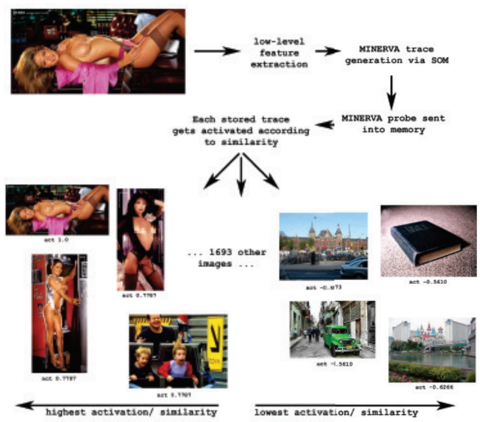

Fig. 2. Example retrieval; one image is entered into the *MINERVA* memory (1702 images/ traces). By using Hintzman's (1988) activation measure, the four best and worst matches are returned.

When building an agent that is supposed to accept low-level input (here, computer images) to simulate complex psychological processes like long-term emotional dynamics, one faces the infamous *symbol grounding problem* voiced by Harnad (1990): How to bridge the gap between sub-symbolic (visual) processing and the handling of discrete/ symbolic (memory) structures?

Cognitive architectures like *ACT-R* (Anderson, 2007) and *CLARION* (Sun, 2006) comprise a complex structure of distinct modules; without offering a seamless integration of emotions and motivations. In contrast, the architecture proposed in Raab et al. (2011) relies on a single memory system inspired by the *MINERVA-2* framework by Hintzman (1988) and enriched by features of the *PSI*-theory (Dörner, 2001) – that integrates motivation and emotion.

Here, we show how images' low-level features (as described by Tamura, 1978) can be used by an unsupervised artificial neural network to produce *MINERVA*-traces (i.e., vectors with *-1*, *0* and *1* as possible values); and that these traces allow for sound memory retrieval – including similarity judgements – at the heart of our agent *PsiCasso* (separate box).

Method

Stimuli: We compiled two sets of images: 1) a set of 1024 images from the web, paying attention to high resolution (at least 1000 px at the short side), crisp colors and distinct themes ranging from nature over artifacts and architecture to humans. 2) we used 682 *Playboy* centerfold pictures (low variance in content and style) as validation set.

Set-up: We implemented a self-organizing map (SOM) in JAVA using the ENCOG engine (Heaton, 2010), with pre-compiled histogram data (3 to 9 bins) and

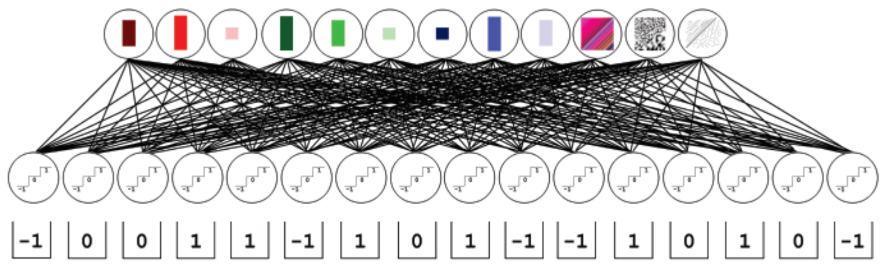

Fig. 1. Color histogram values as well as values for Tamura's contrast, coarseness and directionality enter a self organizing map. Output values are mapped to a *MINERVA* feature vector by calculating the variance of each output neuron and using +/- 1 *SD* for a mapping to 1/-1, 0 otherwise.

Method (Cont.)

Tamura (1978) contrast, coarseness and directionality values (algorithms implemented in JAVA by Nappert, 2010) as input, each feature dedicated to one input neuron. We used our own image set or the Playboy set seperately, and also a combined set – resulting in different output sets. Number of output neurons was varied; the SOM was trained until weights were stable; and each output neuron's value was mapped to a three-step function (-1, 0, 1; see Figure 1). Cut-off value for this discretization was variable using the standard deviation for each neuron's value range. A threshold of 1 *SD* / -1 *SD* resulted in evenly distributed output values.

MINERVA transformation: The resulting vectors were treated as MINERVA-2 vectors (memory model implemented in JAVA, see Neath & Suprenant, 2002), using different combinations of output sets. We tested the quality of MINERVA-2 similiarity retrieval with a variety of images as probe.

Testing: We trained the net using all 1702 images and 25 output neurons. We randomly selected ten centerfolds, probed the memory, and counted how many centerfolds/ people/ other themes were among the 5/ 10 best/ worst matches. We repeated the procedure using ten landscape stills from our collection and counted how many landscapes/ people/ other scenes were found.

Results

Probed with ten centerfolds, on average 65% centerfolds, 7% images with people and 28% other themes were retrieved (considering the ten best matches, results even better for the five most similiar matches). Among the ten least similiar were 9% centerfolds, 6% people and 85% other themes (Figure 3).

Results (Cont.)

For ten landscape probes, retrieved as most similiar were landscapes (42%) in contrast to people (15%) and other scenes (43%). For least similiar: nature (6%), people (43%) and other scenes (51%).

Our results show that an agent equipped with `*MINERVA*-eyes' is fairly good in retrieving similiar images from its memory; even when only a small number of image features is used.

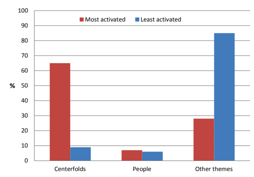

Fig. 3. Similiarity results when memory is probed with ten centerfolds

Moreover, we do not consider the false-positive and – negative retrieval results as *errors*. These mistakes are a kind of fallability that will result in a coupling of images that are similiar (only) in the agent's perception. This way, images that were emotionally neutral in the original perception might get linked to other stimuli – and thus, to emotional states.

Author's note

We wish to thank Ian Neath for providing us with a sample *MINERVA-2* implementation in *JAVA*.

References

- Anderson, J.R. (2007). *How can the human mind occur in the physical universe?* Oxford: Oxford University Press
- Carbon, C. C. (2010). The cycle of preference: Long-term dynamics of aesthetic appreciation. *Acta Psychologica, 134*, 233–244
- Dörner, D. (2001). *Bauplan für eine Seele*. Reinbek: Rowohlt
- Heaton, J. (2010). *Programming Neural Networks with Encog 2 in Java*. St. Louis: Heaton Research Inc.
- Hintzman, D.L. (1988). Judgments of frequency and recognition memory in a multiple-trace memory model. *Psychological Review, 95*(4), 528-551
- Nappert, C. (2010). *The Aesthetics of Photographic Images – Concepts and Approaches for an Automatic Assessment*. Unpublished Bachelor Thesis, University of Bamberg
- Neath, I. and Suprenant, A. (2002). *Human Memory*. Belmont, CA: Wadsworth
- Raab, M., Wernsdorfer, M., Kitzelmann, E. and Schmid, U. (2011). From Sensorimotor Maps to Rules: An Agent Learns from a Stream of Experience. *Lecture Notes in Computer Science, 6830*, 333-339
- Sun, R. (2006). *Cognition and Multi-Agent Interaction*. New York: Cambridge University Press
- Tamura, H., Mori, S. and Yamawaki, T. (1978). Textural Features Corresponding to Visual Perception. *IEEE Transactions on Systems, Man and Cybernetics, SMC-8*(6), 460-473

The *PsiCasso* Framework

Extending the framework by Raab et al. (2011), which is mainly based on a combination of theoretical accounts of Dörner (2001) and Hintzman (1988), our *PsiCasso* agent is based on an emotional-motivational system. Aesthetic appreciation is a form of satisfaction of needs that can take place in two ways:

Direct: Reduction of uncertainty by "understanding" form and content of the visual configuration, e.g. recognizing familiar structures or objects

Associative: Retrospection or anticipation of a satisfaction which is a weakened form of satisfaction

By coupling the neural network described here with a simultaneous perception of the agent's states and needs (also stored in *MINERVA-2-traces*), perception and appreciation become dynamic. Novel input (i.e., low *MINERVA-2* frequency judgement), for example, will challenge the agent's feeling of certainty – and provoke explorative behavior as long as sustainment

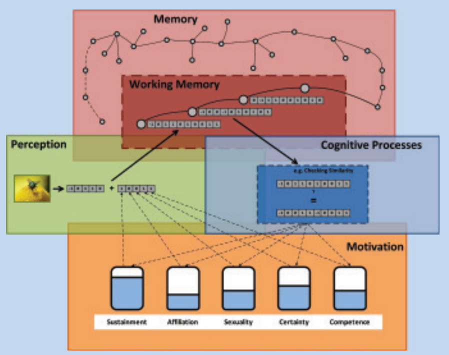

Fig. 4. Design of the *PsiCasso* agent architecture

needs are satisfied. However, after more encounters, the stimulus becomes a source of certainty. In other words: The image appears `boring' to the agent and is devaluated in terms of aesthetics.

Contact: Marius Raab: marius.raab@uni-bamberg.de