Truth or Consequences: Homeostatic Self-Regulation in Artificial Neural Networks

Kingson Man and Antonio Damasio[1]

[1]Brain & Creativity Institute, University of Southern California
kman@usc.edu

Abstract

We assess existing attempts to build emotions and feelings in machines. We review our recently proposed design for machines possessing analogues of biological feeling. Key to our proposal is a homeostatic architecture that regulates internal states to maintain conditions compatible with life. In a first implementation of our design, we present results from a model of synaptic homeostasis in artificial neural networks. We introduce direct consequences to the network's function as a result of its own information processing activity. This model illustrates the benefits that may accrue to a homeostatic learner when it is placed in a needful and vulnerable relation to the objects over which it computes.

Analogues of Feeling in Machines

Living organisms follow the dictates of homeostasis in order to maintain body states in conditions compatible with life. In organisms capable of forming mental states, feelings are the mental expression of these internal viability states. As described in our recent proposal (Man and Damasio 2019), machines that implement a process resembling homeostasis might also acquire a feeling-like device for the motivation and evaluation of behavior.

Our approach diverges from traditional conceptions of intelligence that emphasize outward-directed perception and abstract problem solving. We regard high-level cognition as an outgrowth of resources that originated to solve the ancient biological problem of homeostasis. A physical body is subject to perennial risk and decay. But to a disembodied algorithm—or to a robot whose physical persistence is guaranteed—nothing can happen that carries any consequences.

In our contribution, we first survey recent progress in the neurobiology of conscious feeling and critically assess the existing attempts to build technological analogues of emotions and feelings in machines. Living bodies are composed of living tissues and cells, all subject to perennial risk and to decay in the course of their own regular operations. But nothing equivalent holds for current machines and algorithms whose continuing existence is a given. Organisms do not take a neutral or value-free stance towards information processing. Their actions have consequences for their own well-being. In our perspective, sense-data become meaningful when the data are connected to the maintenance and integrity of the sensing agent and to its homeostasis. Sensory processing that is not attached to a vulnerable body “makes no sense”.

A Model of Synaptic Homeostasis

In a second part, we present results from a model of synaptic homeostasis in artificial neural networks. In biological brains, neurons regulate their intrinsic excitability and synaptic conductance to ensure stable network function (Marder and Goaillard 2006). In artificial neural networks, homeostatic regulation of excitability can reduce saturation and improve signal propagation through the network (Williams and Noble 2007). Such homeostatic regulation of a neural network controller may help to restore behavioral stability under changing conditions (Iizuka and Di Paolo 2008).

In our model, the computing substrate is placed in a needful and vulnerable relation to the very objects over which it computes. Our simulated agents contain convolutional neural networks trained to recognize images of ones and zeroes from the MNIST dataset. Critically, following recognition of an object, the agent must decide to either take it or leave it, that is, it can choose to "ingest" or reject the recognized digit. Feeding on digits helps to replenish the agent's "energy". Critically, the digits also have direct effects on the agent's nervous system — there is no blood-brain barrier here — with ones having excitatory effects and zeroes having inhibitory effects. We implement such synaptic scaling by, for example, varying the slope of the rectified linear activation function.

The agent must answer a question akin to "How does this object make me feel?" We use a counterfactual decision network to answer a related question, "How would my own functionality be affected by taking or leaving this object?" The network simulates the result of either ingesting the digit or not, and subsequently altering its synaptic scaling or not. The accuracy of the network resulting from each scenario is assessed by testing it from a store of previously experienced digits with their associated labels. Agents choose the course of action that best improves their recognition ability. Our elementary model sets up a "strange loop" (Hofstadter 2007) of causality in which accurate recognition is desirable to the agent itself, in order to guide decisions that regulate its internal states and functionality.

References

Hofstadter, D.R., 2007. I am a strange loop. Basic books.

Iizuka H., Di Paolo E.A. (2008) Extended Homeostatic Adaptation: Improving the Link between Internal and Behavioural Stability. In: Asada M., Hallam J.C.T., Meyer JA., Tani J. (eds) From Animals to Animats 10. SAB 2008. Lecture Notes in Computer Science, vol 5040. Springer, Berlin, Heidelberg

Man, K., & Damasio, A. (2019). Homeostasis and soft robotics in the design of feeling machines. *Nature Machine Intelligence*, 1(10), 446-452.

Marder, E. and Goaillard, J.M., 2006. Variability, compensation and homeostasis in neuron and network function. *Nature Reviews Neuroscience*, 7(7), pp.563-574.

Williams, H. and Noble, J., 2007. Homeostatic plasticity improves signal propagation in continuous-time recurrent neural networks. *Biosystems*, 87(2-3), pp.252-259.