

The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain

Motivation(s)

Sensory physiology have achieved appreciable understanding of how information about the physical world is sensed by the biological system. The remaining fundamental questions still shrouding our apprehension of human intelligence are

- In what form is information stored, or remembered?
- How does information contained in storage, or in memory, influence recognition and behavior?

These questions are currently being addressed by two dominant models: symbolic versus connectionist.

The symbolic model suggests that storage of sensory information is in the form of coded representations or images, with some sort of one-to-one mapping between the sensory stimulus and the stored pattern. The appeal of this hypothesis is its simplicity and similarities to a computer. Unfortunately, this method of information retention requires a systematic comparison of the contents of storage with incoming sensory patterns in order to determine whether the current stimulus has been seen before and the appropriate response from the organism.

The connectionist model puts forward that the images of stimuli may never really be recorded at all, and that the nervous system simply acts as an intricate switching network where retention takes the form of new connections or pathways between centers of activity. Since the stored information takes the form of new connections, a new stimuli will make use of new pathways automatically activating the appropriate response.

Proposed Solution(s)

The author proposes a hypothetical nervous system called a perceptron based on the following physical parameters:

x

The number of excitatory connections per A-unit.

y

The number of inhibitory connections per A-unit.

θ

The expected threshold of an A-unit.

ω

The proportion of R-units to which an A-unit is connected.

N_A

The number of A-units in the system.

N_R

The number of R-units in the system.

N_S

The number of S-units in the system.

The current perceptron does not handle temporal stimulus. Since the precise structure of a biological system is unknown, the accompanying analysis makes use of probability theory. This proposed theory of statistical separability is based on the Hayek/Hebb Synaptic Model. The alternative symbolic logic brain models fail to capture equipotentiality and lack neurological correlates. The author asserts that refining these models can never account for biological intelligence because of a difference in principle.

Evaluation(s)

The different modes of organization of a perceptron can be decomposed into

- logical characteristics (α, β, γ)
- discriminanting response systems (mean versus sum), and $\mu \sum$
- reinforcement systems (monovalent versus bivalent).

In the uncompensated gain -system, an A-unit gains an increment of value for every impulse and holds this gain indefinitely. In the constant feed -system, each source-set is allowed a constant rate of gain; the increment is apportioned among the cells of the source-set in proportion to their activity. In the parasitic gain -system, A-units gain in value at the expense of the inactive units of their source-set so that the total value of the source-set is always constant. The responses of these systems to a stimulus are studied in the context of inactive R-units (predominant phase). The overall performance trend is $\alpha\beta\gamma\gamma > \alpha > \beta$

When the R-units are active (postdominant phase), activity is limited to a single source-set as the other sets are suppressed. To understand how the dominant response affects performance, the author restricts the system to mean-discriminanting versus sum-discriminanting. In the former, the R-units whose inputs have the greatest mean value responds first. The latter gives the advantage to the response whose inputs have the greatest net value. The experiments indicate that and system performance are poorer for the model than the case. However, it makes no difference in performance which model is used for the -system. $\alpha\beta\sum\mu\gamma$

The foregoing analysis uses only positive reinforcement (i.e. monovalent). In a bivalent system, one can give positive and negative feedback, and an A-unit may either gain or lose in value depending on the state of the system. The author mentioned that both paradigms perform similarly, but monovalent systems are more efficient.

These results demonstrate that the perceptron surpass existing learning and behavior theories in terms of parsimony, verifiability, and explanatory power and generality. The perceptron only requires each A-unit to carry a hypothetical variable that adheres to certain functional characteristics. When the perceptron fails to adapt in a new situation, one can be confident that either the architecture or the empirical measurements are wrong. The overall theory enables one to reduce complicated tasks to a supervised learning problem and analyze the relationship between neurological variables and learning curves. V

The limits of a perceptron lie in the area of relative judgement and the abstraction of relationships. What is learnable are tasks like naming the color if the stimulus is on the left and the shape if it is on the right. Statistical separability alone does not provide a sufficient basis for higher order abstractions such as name the object left of the square and indicate the pattern that appeared before the circle.

Future Direction(s)

- How to utilize the activation pathways of a neural network to predict the overall accuracy of the system?
- How to quantify the capacity of a system given that one can quantify how much the new associations degrade the old associations?

Question(s)

- Did the author mean that monovalent systems are more efficient to train?
- Are the A-units in the projection area simply routing the input signals?
- The author quantifies the capacity of a system as the probability that a stimulus which has been associated to one of two responses will retain its proper preference after the system has saturated to a given level. Isn't this just the confusion matrix? Wouldn't this make more sense as an accuracy measure?
- Could the inhibitory connections be interpreted as preventing co-adaptation of feature detectors ([../14/dropout-a-simple-way-to-prevent-neural-networks-from-overfitting.html](http://14/dropout-a-simple-way-to-prevent-neural-networks-from-overfitting.html))?

Analysis

In hindsight, this paper illustrates that it is possible to implement a linear classifier when a linearly separable set of binary characteristics (i.e. features) is known. Note that the perceptron as presented here does not include any hidden layers ([../05/multilayer-feedforward-networks-are-universal-approximators.html](http://05/multilayer-feedforward-networks-are-universal-approximators.html)).

The most significant flaw is that the schematic diagrams of a perceptron are too convoluted compared to the ones in [Ros57]. It tries to illustrate a perceptron operates with random connections and can be daisy chained at the expense of decreased clarity. It is difficult to glean from the latest illustration that a response set consists of mutually exclusive R-units.

The author's specification of communality goes against the current trend of learning the right features. Although the system analysis is interesting, it is not obvious how it will extend to multilayer perceptrons.

The most interesting and valuable idea (but not covered in these papers) is the perceptron convergence theorem [MR13]. Note that the perceptron algorithm is described vaguely by [Ros58] as

$$\min \frac{1}{N} \sum_i \max \{0, -y_i \mathbf{w} \cdot \mathbf{x}_i\}$$

where \mathbf{x}_i are the desired pairs of input and output, and \mathbf{w} denotes a separating hyperplane. Notice that the gradient of this hinge loss corresponds to the proposed perceptron learning rule $\{\mathbf{x}_i \in \mathbb{R}^n, y_i \in \{-1, 1\}\}_{i=1}^N, \mathbf{w} \in \mathbb{R}^n$

$$\mathbf{w}_{t+1} = \mathbf{w}_t - \nabla_{\mathbf{w}}.$$

Notes

The Hayek/Hebb Synaptic Model

1. The physical connections of the nervous system which are involved in learning and recognition are not identical from one organism to another. At birth, the construction of the most important networks is largely random, subject to a minimum number of genetic constraints.
2. The original system of connected cells is capable of a certain amount of plasticity; after a period of neural activity, the probability that a stimulus applied to one set of cells will cause a response in some other set is likely to change, due to some relatively long-lasting changes in the neurons themselves.
3. Through exposure to a large sample of stimuli, those which are most similar will tend to form pathways to the same sets of responding cells. Those which are markedly dissimilar will tend to develop connections to different sets of responding cells.
4. The application of positive and/or negative reinforcement may facilitate or hinder whatever formation of connections is currently in progress.

5. Similarity is represented at some level of the nervous system by a tendency of similar stimuli to activate the same sets of cells. It depends on the physical organization of the perceiving system. The structure of the constantly evolving system will determine the classes of things into which the perceptual world is divided.

Organization of a Perceptron

Sensory Units (S-points)

Respond to stimuli on an all-or-nothing basis, or with a pulse amplitude/frequency proportional to the stimulus intensity. Impulses are transmitted to a set of A-units. Any particular S-A connection may be either positive (excitatory) or negative (inhibitory).

Association Units (A-units)

Outputs a value to one or more R-units based on its history and logical characteristics. The set of S-points transmitting impulses to a particular A-unit will be called the origin points of that A-unit.

Response Units (R-units)

Fires according to the McCulloch-Pitts linear threshold unit whose activation function is either mean-discriminating or sum-discriminating. Each R-unit feeds back impulses which inhibit the activity of all mutually exclusive R-units and the A-units that might activate them.

In contrast to the illustration of the perceptron in [Ros57], [Ros58] introduces additional A-units in the form of a projection area. It is important to realize that the A-units in the projection area are not hidden units; they merely implement the switching functions between source and sink.

Notice that the inhibitory connections from the S-points are necessary because a large, complex pattern would always activate any of the A-units which might be activated by its component sub-patterns. Including S-point inhibitory guarantees that the set of A-units responding to a part will generally contain some members which will be inhibited when the whole is presented.

The communality (overlap) between different sets of A-units leads to a statistical interaction between the associations formed. Although this may result in interference between associations, the system gains the ability of combinatorial representations.

References

[MR13] Mehryar Mohri and Afshin Rostamizadeh. Perceptron mistake bounds. *arXiv preprint arXiv:1305.0208*, 2013.

[Ros57(1,2)] Frank Rosenblatt. The perceptron: a perceiving and recognizing automaton. Technical Report, Cornell Aeronautical Laboratory, 1957.

[Ros58(1,2)] Frank Rosenblatt. The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological review*, 65(6):386, 1958.

Source ([../_sources/blog/2016/11/26/the-perceptron-a-probabilistic-model-for-information-storage-and-organization-in-the-brain.rst.txt](#))

[Back to top](#)

© Copyright 2013-2020, alphaXomega.

Created using Sphinx (<http://sphinx-doc.org/>) 2.1.2.