Metadata of the chapter that will be visualized online

Chapter Title	Connectionist Theories of Learning
Copyright Year	2011
Copyright Holder	Springer Science + Business Media, LLC

Corresponding Author	Family Name	**Karaminis**
	Particle	
	Given Name	**Themis N.**
	Suffix	
	Division	Department of Psychological Sciences
	Organization	Birkbeck College, University of London
	Address	Malet Street, London, WC1E 7HX, UK
	Email	tkaram01@students.bbk.ac.uk
Author	Family Name	**Thomas**
	Particle	
	Given Name	**Michael S. C.**
	Suffix	
	Division	Department of Psychological Sciences
	Organization	Birkbeck College, University of London
	Address	Malet Street, London, WC1E 7HX, UK
	Email	m.thomas@bbk.ac.uk

C

Connectionist Theories of Learning

Themis N. Karaminis, Michael S. C. Thomas
Department of Psychological Sciences, Birkbeck College, University of London, London, UK

Synonyms

Associative learning; Backpropagation of error algorithm; Correlational learning; Hebbian learning; Self-organizing maps

Definition

The majority or the connectionist theories of learning are based on the *Hebbian Learning Rule* (Hebb 1949). According to this rule, connections between neurons presenting correlated activity are strengthened. Connectionist theories of learning are essentially abstract implementations of general features of brain plasticity in architectures of artificial neural networks.

Theoretical Background

Connectionism provides a framework (Rumelhart et al. 1986a) for the study of cognition using Artificial Neural Network models. Neural network models are architectures of simple processing units (artificial neurons) interconnected via weighted connections. An artificial neuron functions as a detector, which produces an output activation value determined by the level of the total input activation and an activation function. As a result, when a neural network is exposed to an environment, encoded as activation patterns in the input units of the network, it responds with activation patterns across the units.

In the connectionist framework an artificial neural network model depicts cognition when it is able to respond to its environment with meaningful activation patterns. This can be achieved by modifications of the values of the connection weights, so as to regulate the activation patterns in the network appropriately. Therefore, connectionism suggests that learning involves the shaping of the connection weights. A learning algorithm is necessary to determine the changes in the weight values by which the network can acquire domain-appropriate input-output mappings.

The idea that learning in artificial neural networks should entail changes in the weight values was based on observations of neuropsychologist Donald Hebb on biological neural systems. Hebb (1949) proposed his *cell assembly theory* also known as *Hebb's rule* or *Hebb's postulate*:

> When an axon of cell A is near enough to excite a cell B and repeatedly or persistently takes part in firing it, some growth process or metabolic change takes place in one or both cells such that A's efficiency, as one of the cells firing B, is increased. (1949, p.62)

Hebb's rule suggested that connections between neurons which present correlated activity should be strengthened. This type of learning was also termed *correlational* or *associative* learning.

A simple mathematical formulation of the Hebbian learning rule is:

$$\Delta W_{ij} = \eta a_i a_j \tag{1}$$

The change of the weight (Δw_{ij}) from a sending unit *j* to a receiving unit *i* should be equal to the constant η multiplied by the product of output activation values (α_i and α_j) of the units. The constant η is known as learning rate.

Important Scientific Research and Open Questions

Different learning algorithms have been proposed to implement learning in artificial neural networks. These algorithms could be considered as variants of the Hebbian rule, adjusted to different architectures and different training methods.

A large class of neural networks models uses a multilayered feed-forward architecture. This class of models is trained with *supervised learning* (Fig. 1). The environment is presented as pairs of input patterns and desired output patterns (or targets), where the target is provided by an external system (the notional "supervisor"). The network is trained on the task of producing the corresponding targets in the output when an input pattern is presented.

Norbert Seel (ed.), *Encyclopedia of the Sciences of Learning*, DOI 10.1007/978-1-4419-1428-6,

Comp. by: THAMIZHVEL V Stage: Proof Chapter No.: 398 Title Name: ESL
Page Number: 0 Date:19/4/11 Time:06:45:53

The *Backpropagation of Error* algorithm (Rumelhart et al. 1986b) as proposed for training such networks. Backpropagation is an error-driven algorithm. The aim of the weight changes is the minimization of the output error of the network. The Backpropagation algorithm is based on the *delta rule*:

$$\Delta W_{ij} = \eta(t_i - a_i)a_j \tag{2}$$

The delta rule is a modification of the Hebbian learning rule (Eq. 1) for neurons that learn with supervised learning. In the delta rule, the weight change (Δw_{ij}) is proportional to the difference between the target output (t_i) and the output activation of the receiving neuron (α_i), and the output activation of the sending neuron (α_j).

Backpropagation generalizes the delta rule in networks with hidden layers, as a target activation value is not available for the neurons on these internal layers. Internal layers are necessary to improve the computational power of the learning system. In a forward pass, the Backpropagation algorithm calculates the activations of the units of the network. Next, in a backward pass the algorithm iteratively computes error signals (*delta terms*) for the units of the deeper layers of the network. The error signals express the contribution of each unit to the overall error of the network. They are computed based on the derivatives of the error function. Error signals determine changes in the weights which minimize the overall network error. The *generalized delta rule* is used for this purpose:

$$\Delta W_{ij} = \eta \delta_i a_j \tag{3}$$

According to this rule, weight changes equal to the learning rate times the product of the output activation of the sending unit (α_j) and the delta term of the receiving unit (δ_{ii}).

Although the Backpropagation algorithm has been widely used, it employs features which are biologically implausible. For example, it is implausible that error signals are calculated and transmitted between the neurons. However, it has been argued that since forward projections between neurons are often matched by backward projections permitting bidirectional signaling, the backward projections may allow the implementation of the abstract idea of the backpropagation of error.

Pursuing this idea, other learning algorithms have been proposed to implement error-driven learning in a more biologically plausible way. The *Contrastive Hebbian Learning* algorithm (Hinton 1989) is a learning algorithm for bidirectional connected networks. This algorithm considers two phases of training in each presentation of an input pattern. In the first one, known as the *minus phase* or *anti-Hebbian update*, the network is allowed to settle as an input pattern is presented to the network while the output units are free to adopt any activation state. These activations serve as *noise*. In the second phase (*plus phase* or *Hebbian update*), the network settles as the input is presented while the output units are clamped to the target outputs. These activations serve as *signal*. The weight change is proportional to the difference between the products of the activations of the sending and the receiving units in the two phases, so that the changes reinforce signal and reduce noise:

$$\Delta W_{ij} = \eta\left(a_i{}^{+} a_j{}^{+} - a_i{}^{-} a_j{}^{-}\right) \tag{4}$$

Learning is based on contrasting the two phases, hence the term Contrastive Hebbian Learning.

O'Reilly and Munakata (2000) proposed the LEABRA (Local, Error-driven and Associative, Biologically Realistic Algorithm) algorithm. This algorithm combines error-driven and Hebbian Learning, exploiting bidirectional connectivity to allow the propagation of error signals in a biologically plausible fashion.

The supervised learning algorithms assume a very detailed error signal telling each output how it should be responding. Other algorithms have been developed that assume less detailed information. These approaches are referred to as *reinforcement learning*.

Another class of neural networks is trained with *unsupervised learning*. In this type of learning, the network is presented with different input patterns. The aim of the network is to form its own internal representations which reflect regularities in the input patterns.

The Self-Organizing Map (SOM; Kohonen 1984) is an example of a neural network architecture that is trained with unsupervised learning. As shown in Fig. 2, a SOM consists of an *array of neurons* or *nodes*. Each node has coordinates on the map and is associated with a weight vector, of the same dimensionality as the input patterns. For example, if there are three dimensions in the input, there will be three input units, and each output unit will have a vector of three weights connected to those input units.

The aim of the SOM learning algorithm is to produce a topographic map that reflects regularities in the set of input patterns. When an input pattern is presented to the network, the SOM training algorithm computes the Euclidean distance between the weight vector and the input pattern for each node. The node that presents the least Euclidean distance (*winning node* or *best matching unit [BMU]*) is associated with the input pattern. Next, the weights vectors of the neighboring nodes are changed so as to become more similar to the weights vector of the winning node. The extent of the weight changes for each of the neighboring nodes is determined by its location on the map using a *neighborhood function*. In effect, regions of the output layer compete to represent the input patterns, and regional organization is enforced by short-range excitatory and long range inhibitory connections within the output layer. SOMs are thought to capture aspects of the organization of sensory input in the cerebral cortex. Hebbian learning to associate sensory and motor topographic maps then provides the basis for a system that learns to generate adaptive behavior in an environment.

Cross-References

- ► Adaptive Learning Systems
- ► Associative Learning
- ► Bayesian Learning
- ► Computational Models of Human Learning
- ► Connectionism
- ► Human Cognition and Learning
- ► Learning in Artificial Neural Networks
- ► Parallel Distributed Processing
- ► Reinforcement Learning in Spiking Neural Networks
- ► Self-Organized Learning

References

Hebb, D. O. (1949). *The organization of behavior: A neuropsychological approach*. NewYork: Wiley.

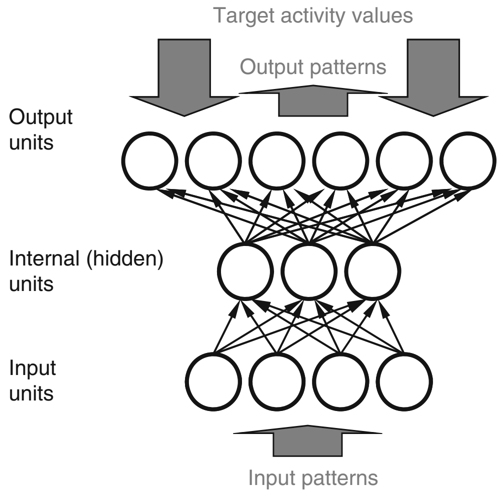
Hinton, G. E. (1989). Deterministic Boltzmann learning performs steepest descent in weightspace. *Neural Computation, 1*, 143–150.

Kohonen, T. (1984). *Self-organization and associative memory*. Berlin: Springer-Verlag.

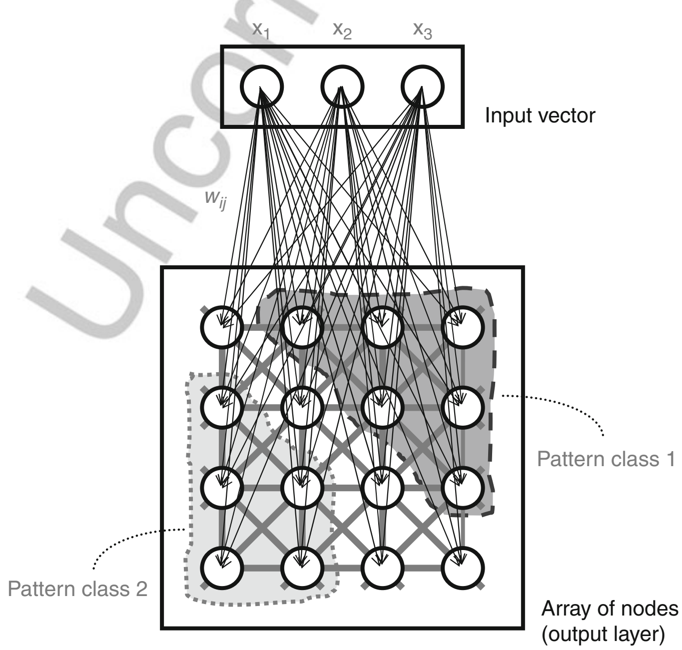
O'Reilly, R. C., & Munakata, Y. (2000). *Computational explorations in cognitive neuroscience: Understanding the mind by simulating the brain*. Cambridge, MA: MIT Press.

Rumelhart, D. E., Hinton, G. E., & McClelland, J. L. (1986a). A general framework for parallel distributed processing. In D. E. Rumelhart, J. L. McClelland, & The PDP Research Group (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition. Volume 1: Foundations* (pp. 45–76). Cambridge, MA: MIT Press.

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986b). Learning internal representations by error propagation. In D. E. Rumelhart, J. L. McClelland, & The PDP Research Group (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition. Volume 1: Foundations* (pp. 318–362). Cambridge, MA: MIT Press.



Connectionist Theories of Learning. Fig. 1 Supervised learning in a three-layered feed-forward neural network



Connectionist Theories of Learning. Fig. 2 Unsupervised learning in a simple self-organizing map (SOM)