Connectionism

Themis N. Karaminis, Michael S.C. Thomas
Department of Psychological Sciences, Birkbeck College, University of London
London, WC1E 7HX
UK
tkaram01@students.bbk.ac.uk, m.thomas@bbk.ac.uk,
http://www.psyc.bbk.ac.uk/research/DNL/

Synonyms

Connectionist Modeling; (Artificial) Neural Network Modeling; Parallel Distributed Processing (PDP); Neural Nets

Definition

Connectionism is an interdisciplinary approach to the study of cognition that integrates elements from the fields of Artificial Intelligence, Neuroscience, Cognitive Psychology, and Philosophy of Mind. As a theoretical movement in Cognitive Science, Connectionism suggests that cognitive phenomena can be explained with respect to a set of *general* information-processing principles, known as Parallel Distributed Processing (Rumelhart, Hinton and McClelland, 1986). From a methodological point of view, Connectionism is a framework for studying cognitive phenomena using architectures of simple processing units interconnected via weighted connections.

These architectures present analogies to biological neural systems and are referred to as *(Artificial) Neural Networks*. Connectionist studies typically propose and implement neural network models to explain various aspects of cognition. The term Connectionism stems from the proposal that cognition emerges in neural network models as a product of a learning process which shapes the values of the weighted connections. Connectionism supports the idea that knowledge is represented in the weights of the connections between the processing units in a distributed fashion. This means that knowledge is encoded in the structure of the processing system, in contrast to the Symbolic approach where knowledge is readily shifted between different memory registers.

Theoretical Background

Artificial Neural Networks are abstract models of biological neural systems. They consist of a set of identical processing units which are referred to as *artificial neurons* or *processing units*. Artificial neurons are interconnected via weighted connections.

A great deal of biological complexity is omitted in artificial neural network models. For example, artificial neurons perform the simple function of discriminating between different levels of input activation. The *detector model* of the neuron (figure 1), is a crude approximation of the role of dendrites and synaptic channels in biological neurons. According to this model, each neuron receives a number of inputs from other neurons. The neuron integrates the inputs by computing a weighted sum of sending activation. Based on the value of the total input activation, an activation function (e.g., a threshold function) determines the level of the output activation of the neuron. The output activation is propagated to succeeding neurons

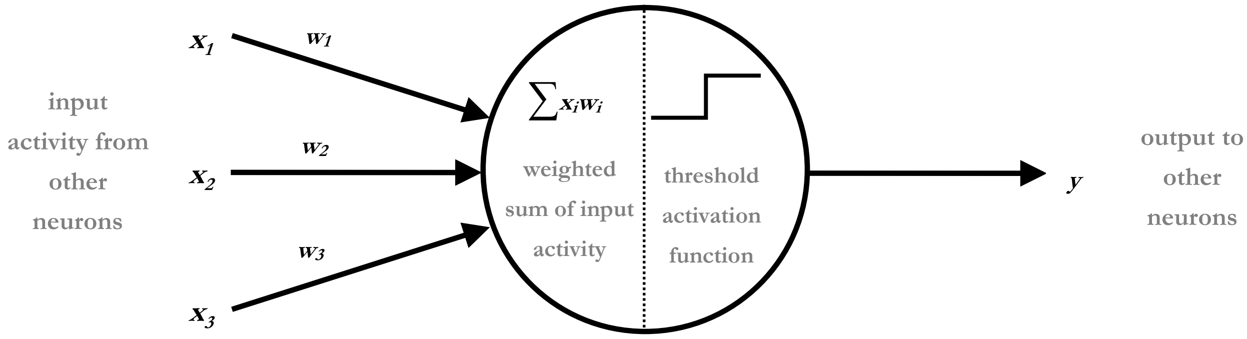
x_1
w_1
input activity from other neurons
x_2
w_2
w_3
x_3
$\sum x_i w_i$
weighted sum of input activity
threshold activation function
y
output to other neurons

Fig. 1. The detector model of the real neuron

The pattern of connectivity between the processing units defines the architecture of the neural network and the input-output functions that can be performed. The processing units are usually arranged in layers. It is notable that a layered structure has also been observed in neural tissues. Many different neural network architectures have been implemented in the connectionist literature. One that has been particularly common is the *three-layer feed-forward neural network* (figure 2). In this network, the units are arranged in three layers: input, hidden and output. The connectivity is feed-forward, which means that the connections are unidirectional, and connect the input to the hidden, and the hidden to the output layer. The connectivity is also full: Every neuron of a given layer is connected to every neuron of the next layer.

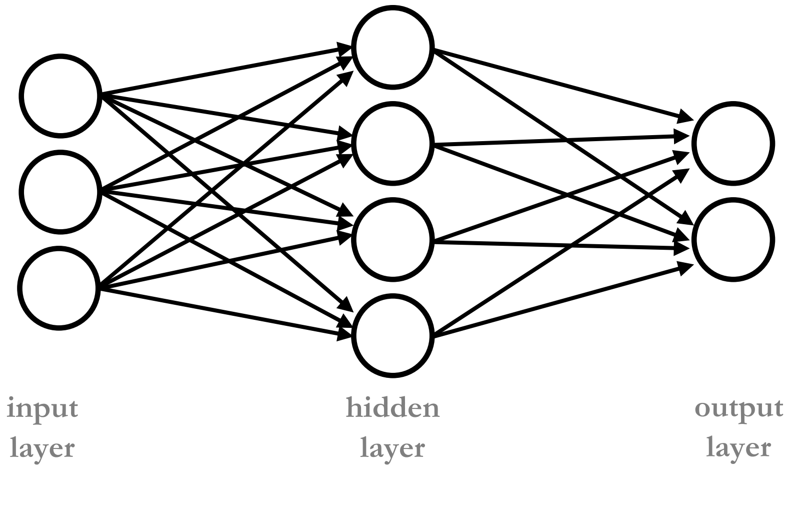

Fig. 2. A three-layered feed-forward neural network with three units in the input layer, four units in the hidden layer, and two units in the output layer.

A key property of neural networks is their ability to learn. Learning in neural networks is based on altering the extent to which a given neuron's activity alters the activity of the neurons to which it is connected. Learning is performed by a *learning algorithm* which determines appropriate changes in the weight values to perform a set of input-output mappings. For example, the *Backpropagation of Error* algorithm (Rumelhart, Hinton, and Williams, 1986) can be used to train a feed-forward multi-layered network (figure 2) using *supervised* learning. For this type of learning, the learning algorithm presents the network with pairs of input patterns and desired output patterns (or targets). The algorithm computes the output error, i.e. the difference between the actual output of the network and the targets. Next, the algorithm propagates appropriate error signals back down through each layer of the network. These error signals are used to determine weight changes necessary to achieve the minimization of the output error. For a more detailed discussion of learning in neural networks, see connectionist theories of learning.

Other issues which are considered in neural network modeling concern the representation of the learning environment. For example, a *localist* or a *distributed* scheme can be used to represent different entities. In the former, a single unit is used to encode an entity, while in the latter an entity is encoded by an activation pattern across multiple units. Furthermore, the different input-output patterns which compose the learning environment can be presented in different ways (e.g., sequentially, randomly with replacement, incrementally, or based on a frequency structure).

Important Scientific Research and Open Questions

The concept of neural network computation was initially proposed in the 1940s. However, the foundations for their systematic application to the exploration of cognition were laid several decades later by the influential volumes of Rumelhart, McClelland, and colleagues. Following this seminal work, a large number of studies proposed neural network models to address various cognitive phenomena.

Although connectionist models are inspired by computation in biological neural systems, they present a high level of abstraction. Therefore, they could not claim biological plausibility. Connectionist models are usually seen as cognitive models, which explain cognition based on general information-processing principles. One of the main strengths of connectionism is that the neural network models are not verbally specified but implemented. In this way, they are able to suggest elaborate mechanistic explanations for the structure of cognition and cognitive development. They also allow the detailed study of developmental disorders by considering training under atypical initial computational constraints, and acquired deficits by introducing 'damage' to trained models.

One of the most influential connectionist models is that of Rumelhart and McClelland (1986) for the acquisition of the English past tense (figure 3). The domain of the English past tense is of theoretical interest to psycholinguists because it presents a predominant regularity, with the great majority of verbs forming their past tenses through a stem-suffixation rule (e.g., walk/walked). However, a significant group of verbs form their past-tenses irregularly (e.g., swim/swam, hit/hit, is/was). Rumelhart and McClelland trained a two-layered feed-forward network (a pattern associator) on mappings between phonological representations of the stems and the corresponding past tense forms of English verbs. Rumelhart and McClelland showed that both regular and irregular inflections could be learned by this network. Furthermore, they argued that their model reproduced a series of well-established phenomena in empirical studies of language acquisition. For example, the past tense rule was generalized to novel stems, while the learning of irregular verbs followed a U-shaped pattern (an initial period of error-free performance succeeded by a period of increased occurrence of *overgeneralization* errors, e.g., *think/thinked* instead of *thought*).

The success of this model in simulating the acquisition of the English past tense demonstrated that an explicit representation of rules is not necessary for the acquisition of morphology. Instead, a rule-like behavior was the product of the statistical properties of input-output mappings. The Rumelhart and McClelland (1986) model posed a serious challenge to existing 'symbolic' views which maintained that the acquisition of morphology was supported by two separate mechanisms, also referred to as the *dual-route model.* According to the dual-route model, a *rule-based system* was involved in the learning of regular mappings, while a *rote-memory* was involved in the learning of irregular mappings. A vigorous debate, also known as the 'past tense debate', ensued in the field of language acquisition (c.f., Pinker & Prince, 1988). By the time this debate resided, connectionist studies had moved on to addressing many aspects of the acquisition of past tense and inflectional morphology in greater detail. For example, Thomas and Karmiloff-Smith (2003) incorporated

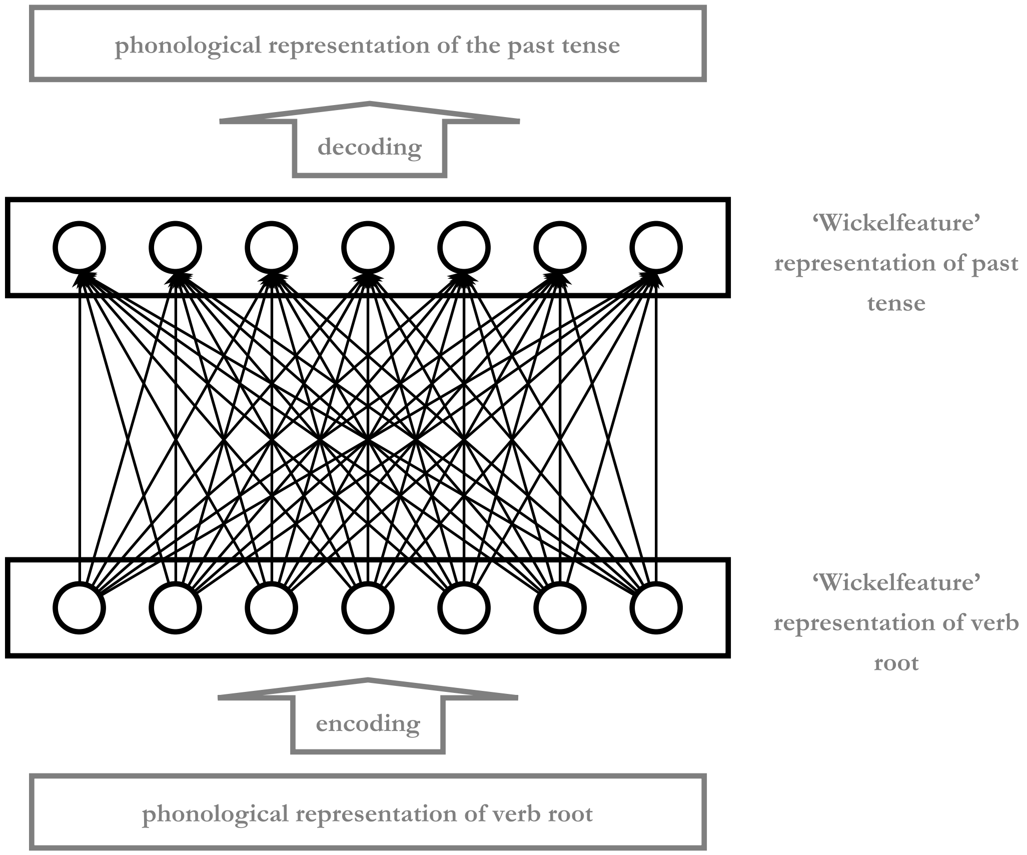
phonological representation of the past tense
decoding
'Wickelfeature' representation of past tense
'Wickelfeature' representation of verb root
encoding
phonological representation of verb root

Fig. 3. The Rumelhart and McClelland (1986) model for the learning of the English past tense. The core of the model is a two-layered feed-forward network (pattern associator) which learns mappings between coarse-coded distributed representations (Wickelfeature representations) of verb roots and past tense forms.

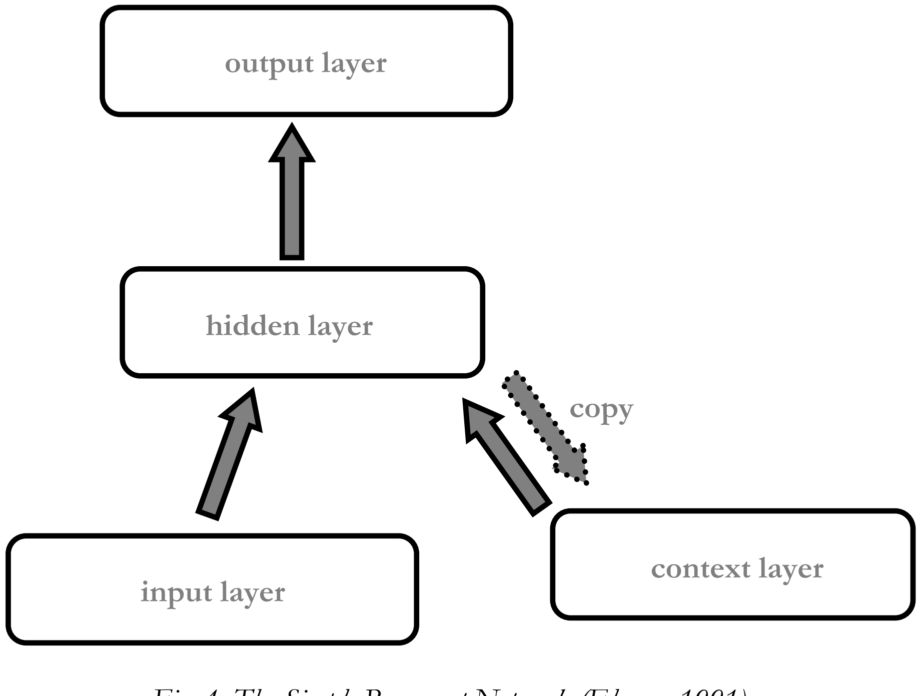

Fig. 4. The Simple Recurrent Network (Elman, 1991).

phonological and lexical-semantics information in the input of a three-layered feed-forward network and studied conditions under which an atypical developmental profile could be reproduced, as a way of investigating the potential cause of developmental language impairments.

Another important connectionist model is the Simple Recurrent Network (SRN) proposed by Elman (1990). The significance of this network lies in its ability to represent time and address problems which involve the processing of sequences. As shown in Figure 4, the SRN uses a three-layered feed-forward architecture in which an additional layer of 'context units' is connected to the hidden layer with recurrent connections. Time is separated into discrete slices. On each subsequent time slice, activation from the hidden layer in the previous time slice is given as input to the network via the context layer. In this way, SRN is able to process a new input in the context of the full history of the previous inputs. This allows the network to learn statistical relationships across sequences in the input.

Acknowledgements

The studies of the first author are funded by the Greek State Scholarship Foundation (IKY). The work of the second author is supported by UK MRC Grant G0300188.

Cross-References

→Connectionist theories of learning

→Learning in artificial neural networks

→Computational models of human learning

→Association learning

→Developmental cognitive neuroscience and learning

→Human cognition and learning

References

Elman, J. L. (1990). Finding structure in time. *Cognitive Science, 14*, 179–211.

Rumelhart, D. E., Hinton, G. E., & McClelland, J. L. (1986). A general framework for parallel distributed processing. In D. E. Rumelhart, J. L. McClelland, & the PDP Research Group, *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. (1986). Learning internal representations by error propagation. In D. E. Rumelhart, J. L. McClelland and The PDP Research Group, *Parallel distributed processing: Explorations in the microstructure of cognition. Volume1: Foundations* (pp. 318–362). Cambridge, MA: MIT Press.

Rumelhart, D. E.,& McClelland, J. L. (1986).On learning the past tense of English verbs. In J. L. McClelland, D. E. Rumelhart & the PDP Research Group (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition, Vol. 2: Psychological and biological models* (pp. 216–271). Cambridge, MA: MIT Press.

Pinker, S., & Prince, A. (1988). On language and connectionism: Analysis of a parallel distributed processing model of language acquisition. *Cognition, 28*, 73–193.

Thomas, M.S.C., & Karmiloff-Smith, A. (2003). Modeling language acquisition in atypical phenotypes. *Psychological Review, 110(4)*, 647–682.