Neural Networks 22 (2009) 568–578

Contents lists available at ScienceDirect

Neural Networks

journal homepage: www.elsevier.com/locate/neunet

2009 Special Issue

A new bidirectional heteroassociative memory encompassing correlational, competitive and topological properties

Sylvain Chartier[a,*], Gyslain Giguère[b,1], Dominic Langlois[a,2]

[a] University of Ottawa, School of Psychology, 200 Lees E217, 125 University Street, Ottawa, ON K1N 6N5, Canada

[b] The University of Texas at Austin, Department of Psychology, 1 University Station, A8000, Austin, TX 78712-0187, USA

ARTICLE INFO

Article history:
Received 6 May 2009
Received in revised form 26 May 2009
Accepted 25 June 2009

Keywords:
Bidirectional heteroassociative memory
Competitive learning
Topological architecture
Unsupervised learning
Categorization
Self-organizing map

ABSTRACT

In this paper, we present a new recurrent bidirectional model that encompasses correlational, competitive and topological model properties. The simultaneous use of many classes of network behaviors allows for the unsupervised learning/categorization of perceptual patterns (through input compression) and the concurrent encoding of proximities in a multidimensional space. All of these operations are achieved within a common learning operation, and using a single set of defining properties. It is shown that the model can learn categories by developing prototype representations strictly from exposition to specific exemplars. Moreover, because the model is recurrent, it can reconstruct perfect outputs from incomplete and noisy patterns. Empirical exploration of the model's properties and performance shows that its ability for adequate clustering stems from: (1) properly distributing connection weights, and (2) producing a weight space with a low dispersion level (or higher density). In addition, since the model uses a sparse representation (k-winners), the size of topological neighborhood can be fixed, and no longer requires a decrease through time as was the case with classic self-organizing feature maps. Since the model's learning and transmission parameters are independent from learning trials, the model can develop stable fixed points in a constrained topological architecture, while being flexible enough to learn novel patterns.

1. Introduction

1.1. Cognitive science background

Every day, humans are exposed to situations in which they are required to either differentiate or regroup perceptual patterns (such as objects presented to the visual system). Their perceptual/cognitive system achieves these operations in order to produce appropriate responses, upon the identity and properties of the encountered stimuli. To accomplish these perceptual tasks, the system must create and enrich context-dependent memory representations, which are adapted to different environments, but can be shared between these environments through generalization. This general process, known as perceptual learning, mainly consists in the implicit abstraction of previously unavailable information, leading to semi-permanent changes at the memory structure level (Gibson & Gibson, 1955). Most perceptual learning processes can be achieved autonomously, through the associative abstraction of the environment's statistical structures (as some neural networks do: Goldstone, 1998; Hall, 1991).

One of the system's main goals in defining mental representations is cognitive economy (Goldstone & Kersten, 2003). Invariance is a quality that leads to economy by reducing the quantity of information that the system must take into account in a given situation, thus accelerating cognitive processes (or more precisely, retrieval of memory traces). Hence, in order to be efficient, the human perceptual/cognitive system must create representations including relevant statistical properties leading to quick differentiation, identification and recognition (Goldstone, 1998). These representations will be useful in future situations, when the need for a decision based on a new or repeating perceptual stimulation arises. Strictly memorizing invariant information lessens the computational burden on the cognitive system, and allows it to follow an information reduction strategy. At the object level, this strategy would be found in the form of a dimensional reduction applied to inputs (Edelman & Intrator, 1997).

Another strategy used by the system to reduce information and increase processing speed is the representation of similar perceptual stimulations by a single higher-level entity, created according to the common perceptual properties of the member objects. This process is called perceptual category learning (Murphy, 2002). Cognitive scientists have historically argued over the fact

* Corresponding author. Tel.: +1 613 562 5800x4419; fax: +1 613 562 5147.
E-mail addresses: sylvain.chartier@uottawa.ca (S. Chartier), giguere.gyslain@mail.utexas.edu (G. Giguère), dlang074@uottawa.ca (D. Langlois).

1 Tel.: +1 512 471 4253.

2 Tel.: +1 613 562 5800x4481.

0893-6080/$ – see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.neunet.2009.06.011

that the human cognitive system either uses generic abstractions (such as prototypes) or very specific perceptual stimulations (such as complete exemplars) to achieve category learning and further classification (Komatsu, 1992). Prototype supporters believe that categories possess a special representational status within the cognitive system. Hence, every exemplar set is linked to a more generic, summary representation that can be retrieved independently, and for which the memory trace is stronger and more durable than for any exemplar (Posner & Keele, 1968, 1970). These representations help the system in quickly deciding on a stimulus' category membership, and are enhanced according to new incoming information. To achieve this, categories must be easy to differentiate; this is why the natural world is composed of cohesive categories, in which associated exemplars are similar to each other (Rosch, 1973).

Following the work of Knowlton and Squire (1993) and Knowlton, Mangels, and Squire (1996), we now know that the prototype theory is valid in categorization. Neuropsychological dissociation studies have led to the conclusion that while single object representations must be memorized in order to achieve recognition, identification and discrimination, categorical representations are completely separate in the system, and operate according to a prototype principle.

1.2. Modeling background

1.2.1. Recurrent/Bidirectional associative memories

When trying to achieve autonomous (or unsupervised) learning and categorization, many neural network options are available. A class of networks known to achieve these types of tasks is that of recurrent autoassociative memories (RAMs). In psychology, AI and engineering, autoassociative memory models are widely used to store correlated patterns. One characteristic of RAMs is the use of a feedback loop, which allows for generalization to new patterns, noise filtering and pattern completion, among other uses (Hopfield, 1982). Feedback enables a given network to shift progressively from an initial pattern towards an invariable state (namely, an attractor).

A problem with these models is that contrarily to what is found in real-world situations, they store learned information using noise-free versions of the input patterns. In comparison, to overcome the possibly infinite number of stimuli stemming, for instance, from multiple perceptual events, humans must regroup these unique inputs into a finite number of stimuli or categories. Also, while RAMs can perfectly reconstruct learned patterns through an iterative process, they cannot associate many distinct inputs with a single representation, that is they cannot categorize.

Direct generalization of RAM models is the development of bidirectional associative memory (BAM) models (Kosko, 1988). BAMs can associate any two data vectors of equal or different lengths (representing for example, a visual input and a prototype or category). These networks possess the advantage of being both autoassociative and heteroassociative memories (Kosko, 1988) and therefore encompass both unsupervised and supervised learning. A BAM model would be able to develop prototype representations linked to different exemplars, but only in a supervised fashion. Nowadays, many RAM/BAM models can display both stability and plasticity for a data set of exemplars (e.g. Davey and Hunt (2000) and Du, Chen, Yuan, and Zhang (2005)).

1.2.2. Competitive networks

Overall, category formation in a perceptual framework is often seen as a process akin to classic clustering techniques, which involve partitioning stimulus spaces in a number of finite sets, or clusters (Ashby & Waldron, 1999). In cognitive modeling, competitive networks (e.g. Grossberg (1988) and Kohonen (1982)) are known for their capacity to achieve clustering behavior. In fact, they constitute local, dynamic versions of clustering algorithms. In these models, each output unit represents a specific cluster. When taking decisions, the association between an exemplar and its determined cluster unit in the output layer is strengthened. In *winner-take-all* (WTA) networks (Grossberg, 1988; Kohonen, 1982), exemplars may only be associated with one cluster (i.e. only one output unit at a time can be activated).

An example of one such *hard competitive* framework is that of the Adaptive Resonance Theory (ART: (Grossberg, 1988)). ART networks possess the advantage of being able to deal effectively with the exemplars/prototype scheme, while solving the stability/plasticity dilemma. These unsupervised models achieve the desired behavior through the use of a novelty detector (using *vigilance*). Various degrees of generalization can be achieved by this procedure: low vigilance parameter values lead to the creation of broad categories, while high values lead to narrow categories, with the network ultimately performing exemplar learning.

Another example of this framework is the *self-organizing feature map* (SOFM: Kohonen, 1982), which, in addition to showing WTA behaviors, uses a topological representation of inputs and outputs. Although SOFMs only consider one active output unit at a time, the learning algorithm also allows for physically close neighboring units to update their connection weights. In a SOFM, an exemplar may thus, for instance, be geometrically positioned between two clusters, and possess various degrees of membership.

An extension of the WTA principle selects the k largest outputs from the total n outputs (Majani, Erlanson, & Abu-Mostafa, 1989). This *k-winners-take-all* (kWTA) rule is thus a more general case of the WTA principle, within which exemplars may be associated with many clusters at differing degrees. This procedure provides a more distributed classification.

While extremely useful for object and category processing, competitive networks do not achieve recurrent behavior. They are thus generally sensitive to input noise during recall. ART networks can deal with noise through a novelty detection procedure, but do not encompass topological properties. In comparison, SOFMs show topological properties, but are not resistant to noise and do not show plasticity.

1.3. Goals and presentation

In the present paper, we propose a new bidirectional heteroassociative memory (BHM), named BHM + SOFM, which encompasses kWTA and SOFM properties. This modification will enable a known BHM model (Chartier & Boukadoum, 2006a) to increase its clustering capacity; hence, higher network performance and readability should follow. In addition, this kWTA–SOFM version of a BHM will allow for sparse coding, which is a distributed representation principle supported by neuropsychological findings (Olshausen & Field, 2004). The network will also inherit properties from its recurrent memory status, such as attractor development and noise tolerance (Hassoun, 1989). Finally, using sparse coding, we will propose a modification to the original network (Chartier, Giguère, Langlois, & Sioufi, in press) that will enable it to solve a simple version (exemplar data set) of the stability–plasticity problem (Grossberg, 1987). Therefore, the goal of this study is to propose a general model based on BAMs, that shows properties similar to those found in other network types, such as competitive behavior in SOFMs.