Hierarchical Temporal Memory in Visual Pattern Recognition

David Rozado
Thesis Advisor: Pablo Varona Martínez

Overview

The present work consisted on the implementation of a set of algorithms designed to carry out visual pattern recognition. The algorithms pulí heavily from a number of principies that are thought to underlie neocortical mechanisms that give rise to sensory perception and higher cognitive functions.

These set of algorithms represent a formalization of a framework known as Hierarchical Temporal Memory (HTM).

HTM was originally proposed by Jeff Hawkings at the Redwood Institute for theoretical Neuroscience. HTM theory is a speculative top-down theory about neocortical function. The neocortex is a fine layer of tissue that is wrapped around and folded over the rest of the brain, becoming the most salient feature of the human brain for an external observer [3]. The neocortex is thought to play a critical (if not main) role in cognitive tasks such as pat-tern recognition during sensory perception, organization of motor behavior, executive control and memory storage and recall [2],

HTM suggests an overall framework about neocortical function by using a number of well known principies about brain function discovered through time by neurophysiologists, neuroanatomists, and theoretical neuroscientists [4]. The HTM theory pavés the way for formalizing this set of principies into a group of algorithms that when implemented on a computer can be used to replícate, in an approximative manner, some of the functions that the human neocortex is able to perform.

State of the art

Vision is the primary sensory modality for humans and most mammals to perceive the world. In humans, visión related áreas occupy about 30 per-cent of the neocortex [5]. Light rays are proyected upon the retina and the brain tries to make sense of the world by means of interpreting this visual input pattern. The sensitivity and specifity with which the brain solves this computationally complex problem can not yet be replicated on a computer. The most imposing of these problems is that of invariant visual pattern recognition. Humans and mammals in general can recognize images despite changes in location, size, ligthting conditions and in the presence of defor-mations and large amounts of noise. Several approaches have been used to solve this problems, usually consisting on having a set of labeled examples from a certain number of categories of objects. Most approaches either pre-program or learn low-level statistical patterns("features") of the images, then try to map combinations of these features to the correct categories using a simple classifier[6]. Some level of invariance is achieved when trained with tranlations, deformations, and size changes of the original objects. Nonethe-less, these techniques experience difficulties when trying to generalize from patterns to which the system has been exposed to novel patterns.

There are two cues to how the neocortex might solve the invariance recog-nition problem that have largely been ignored in the computer science litera-ture. A lot of research has not payed attention to the role that time may play in mammalian visión as well as to the fact that most of the learning carried out by mammals during the development of visual pattern recognition skills is inherently unsupervised. HTM conciles this two cues by means of inte-grating information over múltiple time steps and using time as a supervisor to the systemfl].

Methodology

To test our algorithms, we proyected simple Une drawing corresponding to 48 different categories of objects over an artificial retina of size 32 by 32 pixels and move them horizontally and vertically using several scaled versions of each object over the entire visual field. There were 5 exemplar images of different scales for each category. The purpose of the system was to be able to recognize, after training, images from the set of training images subjected during inference to distortion and noise.

We created a hierarchical system of 3 layers, each layer was composed of an array of several nodes that were trained using the movies described in the previous paragraph. The lowest level, level 1, consisted of modules receiving inputs from a 4x4 patch of the input image. Sixty four level 1 modules tiled an input image. Learning started at level 1 and proceeded to the higher levéis. At level 2, modules received its inputs from 4 adjoining level 1 modules. There were a total of 16 level 2 modules. A single level 3 module received all the information from these level 2 modules, and assign the input patterns to a particular image category.

Basically each node at each different level was set up to store a number of common coincidences (events that happen cióse together in space) and

cluster them into sets of coincidences that are likely to follow each other in time, (events that oceur cióse together in time). Each node therefore forms a memory of non-overlaping invariant representations of patterns to which this node is being exposed to, within its receptive field. Each of these patterns correspond to certain objects or subparts of it. That is, each node pulled together different coincidences that oceured cióse together in time and space and assigned them a common invariant representation at each different level
in the hierarchy.

The fact that the world consists of a hierarchical structure that the system tries to capture, leads to generalization of learning and inference as well as storage efficiency. Time plays a critical role in our system since it func-tions as a de-facto supervisor during learning by means of asigning different coincidences (that may have a large geometric distance from each other) to the same invariant representation as long as they have been presented cióse together in time often enough during training.

Results

The system was able to discriminate among test images from a library of 48 different image categories. Each image category being composed of 5 instances of different sizes. The system was robust to scale, translation and distortion invariance even in the presence of noisy input. The invariances developed (translation, size and distortion) by the system were the ones the system was trained to be invariant to. Theoretically the system could be trained to develop invariance to rotation, but the input space would increase considerably and the computing power required to perform the experiment would increase significantly.

Conclusión

The purpose of this work has been to establish a proof of concept of the HTM theory as well as the formalization of this theory into'a set of in-house developed algorithms. Our aim for the future is to explore the application of this algorithms to a wide array of different problems as well as to explore and expand this algorithms by means of the addition of further -biologically inferred- principies of neocoftical function.

References

[1] Dileep George, Bobby Jarosy, *The HTM Learning Algorithms*, Numenta, Inc. www.numenta.com, (2007).

[2] Jeff Hawkings, Dileep George, *Hierarchical Temporal Memory, Concepts, Theory, and Terminology*, Numenta, Inc. www.numenta.com, (2006).

[3] Vermon B. Mountcastle The columnar organization of the neocortex Hierarchical. [Review]. Brain (1997), 120, 701722

[4] Dileep George, Jeff Hawkings *A* hierarchical Bayesian Model of Invariant Pattern Recognition in the Visual Cortex. Neural Networks, 2005.

IJCNN '05. Proceedings. 1812- 1817 vol. 3

[5] Rodney J. Douglas and Kevan A.C. Martin Neuronal Circuits of the Neocortex (2004). Annual Review of Neuroscience Vol. 27: 419-451.

[6] Christopher M. Bishop. Neural Networks for Pattern,Recognition. Oxford University Press, USA (November 3, 1995)