Image Object Recognition Based on Biologically Inspired Hierarchical Temporal Memory Model and Its Application to the USPS Database

[1]S. Štolc, [2]I. Bajla

[1]Institute of Measurement Science, Slovak Academy of Sciences, Bratislava, Slovakia,
[2]Austrian Research Centers GmbH – ARC, Smart Systems Division, Seibersdorf, Austria
Email: umerstol@savba.sk

Abstract. *In the paper we describe basic functions of a Hierarchical Temporal Memory (HTM) based on a novel biologically inspired network model of the overall large-scale structure of human neocortex. It appeared in a form of research release of the system NuPIC (Numenta Platform for Intelligent Computing) in 2007. In the design of the HTM, hierarchical structure and spatio-temporal relations serve for generation of invariant representations of the outer world (e.g. world of visual patterns), similar to those in human neocortex. There are several open issues for a research into HTM, in particular those applied to pattern recognition tasks. In the paper we report our results of the HTM architecture design and optimization of the network parameters for the task of recognition of the handwritten digits from the well benchmarked USPS database.*

Keywords: Hierarchical Temporal Memory (HTM), Pattern Recognition, USPS

1. Introduction

The HTM is an uncommon, neocortex inspired computational theory [1] that appeared in a form of research release of the software NuPIC 1.6. (Numenta Platform for Intelligent Computing) in 2007 [2]. As an original memory-prediction theory, it has a potential to mature in the future into the state capable to solve also various problems of pattern recognition. In the design of the HTM, hierarchical structure and spatio-temporal relations serve for generation of invariant representations of the world (in our domain restricted to images), similar to those in human neocortex. The functioning of biological regions and subregions is simulated in the nodes (basic units of the HTM) using Bayesian belief revision techniques. The basic difference of the HTM to the neural networks consists in providing a model of the overall large-scale structure of human neocortex. There are several open issues for a research into HTM, especially those applied to specific Pattern Recognition (PR) tasks. In the paper we report our results of HTM architecture design and optimization of the parameters of the individual HTM-nodes for the task of recognition of the handwritten digits of the internationally accepted USPS database. This choice is based on the fact that the recognition accuracy achieved for various classifiers applied to this database is well benchmarked in the literature, and it can be compared to the results of our research into object detection using vector subspace methods, in particular non-negative matrix factorization methods [3].

2. Hierarchical Temporal Memory

In [4,5], the general concept, theory, as well as terminology of the HTM is described. Our interests were focused on research into optimal design of such an HTM that can be used for solving image recognition tasks, therefore we will briefly describe the basic functions of the HTM implementation in relation to modelling the visual world. The HTM is a memory-prediction network that is organized in several layers of elementary units – nodes working in identical mode. The individual layers (levels) are ordered in a hierarchical tree-like structure.

There is a zero sensory level of the HTM which serves as an input to the first level of nodes. In our case zero level (ImageSensor) represents a visual field of image pixels. Since the use of temporal dependences of input spatial patterns is essential characteristic of the HTM, it learns either from natively moving images or sequences of image frames of an artificially generated movie (obtained by applying a limited set of translations, rotations and zoomings to the given training images). It is the explorer plug-ins that generate the temporal sequences within the ImageSensor object. They are responsible for "exploring" the input space of possible images accomplishing two main goals: (i) efficiently select images for presentation to the network, out of a large space of potential images, (ii) generate smooth temporal sequences needed for training TemporalPoolerNode. When the learning process is finished in all levels, the HTM network can classify an unknown pattern into previously defined classes. For every input, the node does three learning operations: (i) memorization of the input vectors, (ii) learning transition probabilities, (iii) temporal grouping.

The memorization of the pattern vectors is carried out in the spatial pooler that actually generates spatial statistical representations of input vectors (patterns). More specifically: at the 1st hierarchy level, the spatial pooler of each node detects clusters of pattern vectors occuring in the field of view of this node in the course of training. Each cluster is memorized by means of a centroid representative. At higher levels of the HTM hierarchy, the spatial clustering algorithm is applied to belief vectors which are input for higher level spatial poolers. Once the memorization process is finished, the spatial pooler can continue with the next step of the learning process. During this stage, for every input pattern its closeness to every vector stored in the node memory [6] is measured by Euclidean distance d_i of the pattern vectors. It is assumed that the probability that the input pattern matches the i-th stored pattern vector can be calculated as being proportional to the Gaussian function $e^{-d_i^2/\sigma^2}$ of the distance d_i, where σ is a parameter of the node.

The learning in the temporal pooler is characterized as follows. First, a time adjacency matrix for the pattern vectors is generated, entries of this matrix are numbers of transition events between vectors following each other during image movement in the field of view of the node (the rows and columns of the adjacency matrix represent memorized pattern vectors – coincidences – of the given node. Second, for temporal grouping of the pattern vectors (level 1) or beliefs (other levels), the Agglomerative Hierarchical Clustering is used. In contrast to the clusters generated in the spatial poolers, these clusters reflect temporal dependences and therefore they are called temporal groups. Each temporal group can be seen as an invariant representation of the patterns included in this group.

The topmost HTM level is constituted by the only node – supervised classifier (in our case Zeta1TopNode) [2] – in which a traditional supervised grouping of a network input is carried out based on the corresponding beliefs produced by preceding HTM levels.

3. Application to USPS database

Train and test sets

For the purpose of testing the performance of the HTM model in comparison with other classification approaches, we have decided on the standard USPS (U.S. Post Service) database of handwritten digits collected by CEDAR, Buffalo [7] and later on converted to gray level format by LeCun's research group [8]. The USPS database consists of 9298 digits of 16x16 pixels each which are divided into two non-overlapping groups: 7291 digits for training and 2007 digits for testing. The most pronounced advantage of using this data set is that a vast number of benchmarks has been performed for different classification methods.

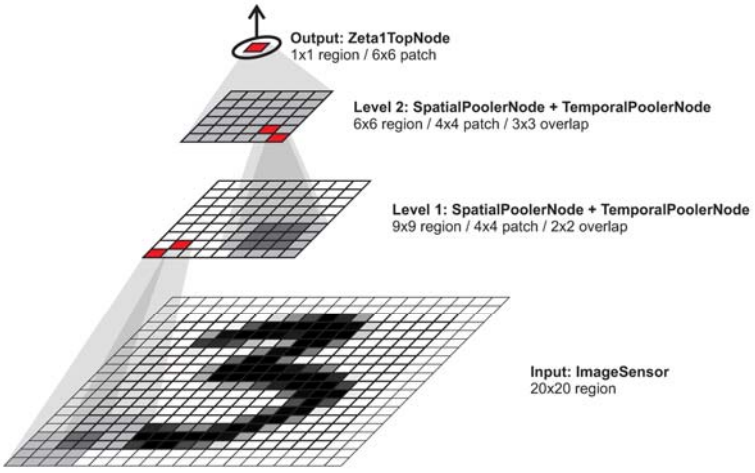

Fig. 1. In the picture 2-level HTM architecture is depicted. The network consists of ImageSensor (input), two levels comprising both SpatialPoolerNode and TemporalPoolerNode regions (level 1 and level 2), and Zeta1TopNode (output).

Optimization of the HTM network architecture

We have experimented with various types of 1, 2 and 3-level architectures where multiple parameters had to be tuned. Let us focus on the specific network architectures which showed the best overall recognition accuracy. The basic assumption originating from the USPS database itself and holding for all our network architectures is that every image passed to the network must have the same size of 16x16 pixels. To avoid undesired border effects, the input images are padded by 2-pixel border while getting the dimension of 20x20 pixels. We have found the following architecture optimal for the given task: a 2-level HTM network that consists of ImageSensor (input), two levels comprising joined SpatialPoolerNode and TemporalPoolerNode regions (level 1 and level 2), and Zeta1TopNode (output). The nodes at level 1 are arranged to a 9x9 array such that they cover full extent of the ImageSensor (20x20 pixels). Each level 1 node receives input from a 4x4 patch while two neighboring patches overlap by 2 nodes. The nodes at level 2 form a 6x6 array while each node receives input again from a 4x4 patch with 3 node overlap. The output Zeta1TopNode receives input from all nodes at level 1 (see Fig. 1).

There are two parameters of each SpatialPoolerNode: *maxDistance* and σ of the Gaussian which are to be optimized at every network level (identical values use all nodes at one level). The TemporalPoolerNodes at every level have only one parameter: *requestedGroupCount* that should be optimized. HTM documentation does not prescribe or include any recommendation or method to be applied for parameter optimization in a particular HTM application. The decision is left on the user. We have decided to use a mixture of two approaches: some of the parameters were estimated using perpendicular-search method [9], whereas other parameters were estimated according to the experience or recommendations from Numenta discussion forum. The perpendicular-search method is based on the principle of searching for the minimum of error function along individual parameters, however, in each iteration the parameter is altered only within the restricted area around the starting point. Best obtained parameter combination is then chosen as the starting point for the next iteration. This means that only one parameter is varied at a time while all other parameters keep the same value.

A key capability of the ImageSensor, which reads data from image files and hands it off to nodes in an HTM network [6], is the ability to generate smoothly-varying patterns forming “virtual” temporal sequences (movies). The ExhaustiveSweep explorer is one of the most

common explorers implemented within NuPIC 1.6.1. This explorer performs an exhaustive raster scan through the input space. It can generate rather complex sequences by translating an image side-to-side either horizontally or vertically always by one pixel. The sequences generated are rather long and therefore imply highly memory and time demanding training phase. Since the USPS database contains normalized patterns only (i.e. numbers are always centered and their size is normalized to 16x16 pixels), it is not necessary to build up a positional and dimensional invariance for this data. The ExhaustiveSweep explorer appeared to be inappropriate in this case mainly due to its high computational costs. We have developed an alternative explorer which better meets our expectations involved by the USPS database. Basic idea, on which this explorer is inspired, is a way of how humans are seeing the letters while reading a text. When eyes are moving through the text lines, each single symbol is being seen from different viewing angles. The symbol in the top-left corner of a page looks slightly different than the same symbol (i.e. same character, font and size) in the center or bottom-right corner of a page. The ViewAngleSweep explorer tries to imitate this behavior by smooth alternating the viewing angle in 9 fixed values which form the final temporal sequence presented to the network.

4. Results

The results of the design of suboptimal HTM network for the application to the USPS handwritten digits database can be summarized as follows. The 2-level architecture, characterized in Fig. 1, has been proposed. The following optimum values of the adjustable network parameters have been found:

- Level 1:
  - SpatialPoolerNode: *maxDistance* = 250, σ = 500;
  - TemporalPoolerNode: *requestedGroupCount* = 80;
- Level 2:
  - SpatialPoolerNode: *maxDistance* = 0.3, σ = 1;
  - TemporalPoolerNode: *requestedGroupCount* = 950.

All the nodes in the individual HTM levels have been learned using a special ViewAngleSweep explorer that we developed. We have achieved the overall classification

Table 1. Confusion matrix for 2-level HTM architecture.

True \ Class	0	1	2	3	4	5	6	7	8	9	Accuracy
0	355	0	2	0	0	0	0	1	1	0	98.86 %
1	0	259	0	0	2	0	2	1	0	0	98.11 %
2	1	0	190	0	1	0	0	3	3	0	95.96 %
3	1	0	0	155	0	6	0	1	2	1	93.37 %
4	0	2	0	0	185	1	1	6	0	5	92.50 %
5	1	0	1	1	0	154	0	0	1	2	96.25 %
6	1	0	0	0	0	1	166	0	2	0	97.65 %
7	0	1	1	0	1	0	0	144	0	0	97.96 %
8	5	0	0	2	0	3	0	0	154	2	92.77 %
9	1	0	0	0	1	1	0	0	0	174	98.31 %

accuracy: 96.46 %. For a detailed report on classification accuracy on USPS testing set see table 1.

5. Conclusions

In the paper we have described a suboptimal design of the HTM network when applied to the task of image recognition, in particular, to handwritten digits of the USPS database. Comparison of the obtained results to the results achieved by other classifiers published in [10] showed that the HTM overcomes performance of a number of tested classifiers, and only several of them achieved higher classification accuracy (range between 97-98 %, combination of tangent vector and local representation and SVM-like approaches). Two final conclusions can be drawn: first, in the HTM, the Zeta1TopNode can be replaced and tuned for application of the SVM classifier, second, as commented by D. George, real power of the HTM architecture can be demonstrated in tasks in which a real temporal hierarchy (instead of a virtual movie) occurs.

Acknowledgements

This work has been supported by the Forschung Austria (grant “Application potential of HTM – a new paradigm of intelligence”).

References

[1] D. George and J. Hawkins. Hierarchical bayesian model of invariant pattern recognition in the visual cortex. In *Proceedings of the international joint conference on neural networks*, Montreal, Canada, 2005. International Neural Networl Society.

[2] Numenta. *Zeta1 Algorithms Reference*, March 2007. Document version 1.0.

[3] D. Soukup and I. Bajla. Robust object recognition under partial occlusions using NMF, 2008. http://www.hindawi.com/getarticle.aspx?doi=10.1155/2008/857453

[4] Numenta. *Hierarchical Temporal Memory, Concepts, Theory, and Terminology*, June 2008. Document version 1.8.0.

[5] Dileep George. *How the Brain Might Work: A Hierarchical and Temporal Model for Learning and Recognition*. PhD thesis, Stanford University, 2008.

[6] Numenta. *Numenta Node Algorithms Guide, NuPIC 1.6*, June 2008.

[7] C. H. Wang and S. N. Srihari. A framework for object recognition in a visually complex environment and its application to locating address blocks on mail pieces. *International Journal of Computer Vision*, 2(2):125–151, September 1988.

[8] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel. Backpropagation applied to handwritten zip code recognition. *Neural Computing*, 1(4):541–551, 1989.

[9] M. S. Bazaraa, H. D. Sherali, and C. M. Shetty. *Nonlinear Programming, Theory and Algorithms*. John Wiley & Sons Inc. New York, 1990. ISBN 0-471-59973-5.

[10] J. Dong. HeroSvm 2.1, 2005. http://www.cenparmi.concordia.ca/~jdong/HeroSvm.html