Sparse Coding in Sparse Winner Networks

Janusz A. Starzyk[1], Yinyin Liu[1], David Vogel[2]

[1] School of Electrical Engineering & Computer Science
Ohio University, Athens, OH 45701
{starzyk, yliu}@bobcat.ent.ohiou.edu[1]
[2] Ross University School of Medicine
Commonwealth of Dominica
dvogel@rossmed.edu.dm

Abstract. This paper investigates a mechanism for reliable generation of sparse code in a sparsely connected, hierarchical, learning memory. Activity reduction is accomplished with local competitions that suppress activities of unselected neurons so that costly global competition is avoided. The learning ability and the memory characteristics of the proposed winner-take-all network and an oligarchy-take-all network are demonstrated using experimental results. The proposed models have the features of a learning memory essential to the development of machine intelligence.

Keywords. sparse coding, winner-takes-all, Hebbian learning in sparse structure, hierarchical self-organizing memory

1 Introduction

In this paper we describe a learning memory built as a hierarchical, self-organizing network in which many neurons activated at lower levels represent detailed features, while very few neurons activated at higher levels represent objects and concepts in the sensory pathway [1]. By recognizing the distinctive features of patterns in a sensory pathway, such a memory may be made to be efficient, fault-tolerant, and to a useful degree, invariant. Lower level features may be related to multiple objects represented at higher levels. Accordingly, the number of neurons increases up the hierarchy with the neurons at lower levels making divergent connections with those on higher levels [2]. This calls to mind the expansion in number of neurons along the human visual pathway (e.g., a million geniculate body neurons drive 200 million V1 neurons [3]).

Self-organization is a critical aspect of the human brain in which learning occurs in an unsupervised way. Presentation of a pattern activates specific neurons in the sensory pathway. Gradually, neuronal activities are reduced at higher levels of the hierarchy, and sparse data representations, usually referred to as “sparse codes”, are built. The idea of “sparse coding” emerged in several earlier works [4][5]. In recent years, various experimental and theoretical studies have supported the assumption that information in real brains is represented by a relatively small number of active neurons out of a large neuronal population [6][[7][3].

In this paper, we implement the novel idea of performing pathway selections in sparse network structures. Self-organization and sparse coding are obtained by means of localized, winner-take-all (WTA) competitions and Hebbian learning. In addition, an oligarchy-take-all (OTA) concept and its mechanism is proposed that produces redundant, fault tolerant, information coding.

This paper is organized as follows. In section 2, a winner network is described that produces sparse coding and activity reduction in the learning memory. In section 3, an OTA network is described that produces unsupervised, self-organizing, learning with distributed information representations. Section 4 demonstrates the learning capabilities of the winner and the OTA networks using experimental results. Finally, our method of sparse coding in sparse structures is summarized in section 5.

2 The Winner Network

In the process of extracting information from data, we expect to predictably reduce neuronal activities at each level of a sensory pathway. Accordingly, a competition is required at each level. In unsupervised learning, we need to find the neuron in the network that has the best match to the input data. In neural networks, such a neuron, is usually determined using a WTA network [8][9]. A WTA network is usually implemented based on competitive neural network in which inhibitory lateral links and recurrent links are utilized, as shown in Fig. 1. The outputs iteratively suppress the signal strength among each other and the neuron with maximum signal strength will stay as the only active neuron when the competition is done. For a large memory, with many neurons on the top level, a global WTA operation is complex, inaccurate and costly. Moreover, average competition time increases as the likelihood of similar signal strengths increases in large WTA networks.

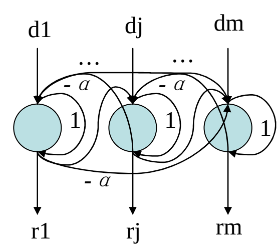

Fig. 1. WTA network as competitive neural network

The use of sparse connections between neurons can, at the same time, improve efficiency and reduce energy consumption. However, sparse connections between neurons on different hierarchical levels may fail to transmit enough information along the hierarchy for reliable feature extraction and pattern recognition. In a local network model for cognition, called an “R-net” [10][11], secondary neurons, with random connections to a fraction of primary neurons in other layers, effectively provide almost complete connectivity between primary neurons pairs. While R-nets provide large capacity, associative memories, they were not used for feature extraction and sparse coding in the original work.

The R-net concept is expanded in this work by using secondary neurons to fully connect primary neurons on lower levels to primary neurons on higher levels through the secondary neurons of a sparsely connected network. The network has an increasing number of neurons on the higher levels, and all neurons on the same level have an equal number of input links from neurons on the lower level. The number of secondary levels between primary levels affects the overall network sparsity. More secondary levels can be used to increase the network sparsity. Such sparsely connected network with secondary levels is defined as winner network and illustrated in Fig. 2.

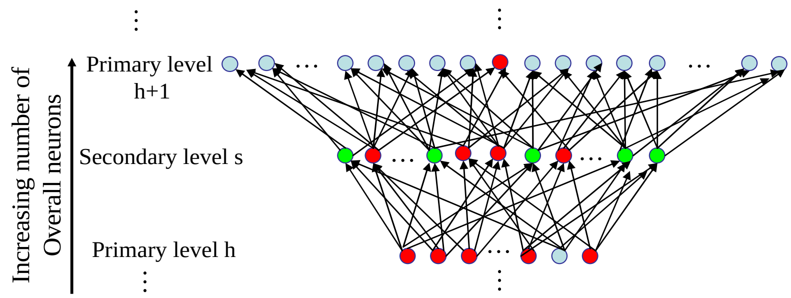

Fig. 2. Primary level and secondary level in winner network

The initial random input weights to each neuron are scaled to have a sum of squared weights equal to 1, which places them on the unit multidimensional sphere. Because a neuron becomes active when its input weight vector is similar to its input pattern, spreading the input weights uniformly on the unit-sphere increases the memory capacity of the winner network. Furthermore, the normalization of the weights maintains the overall input signal level so that the output signal strength of neurons, and accordingly the output of the network, will not be greatly affected by the number of input connections.

In a feed-forward computation, each neuron combines its weighted inputs using a thresholded activation function. Only when the signal strength is higher than the activation threshold can the neuron send a signal to its post-synaptic neurons. Eventually, the neurons on the highest level will have different levels of activation, and the most strongly activated neuron (the global winner) is used to represent the input pattern. In this work, the competition to find the global winner is replaced by small-scale WTA circuits in local regions in the winner network as described next.

In a sparsely connected network, each neuron on the lower level connects to a group of neurons on the next higher level. The winning neuron at this level is found by comparing neuronal activities. In Hebbian learning, weight adjustments reduce the plasticity of the winning neuron’s connections. Therefore, a local winner should not only have the maximum response to the input, but also its connections should be flexible enough to be adjusted towards the input pattern so that the local winner is,

$$s_{winner\ i}^{level+1} = \max_{j \in N_i^{level+1}} \left\{ \sum_{k \in N_j^{level}} w_{jk} s_k^{level} \cdot \rho_{ji} \right\} \quad (i = 1,2,..N^{level}), \tag{1}$$

where $N_i^{level+1}$ is a set of post-synaptic neurons on level *(level+1)* driven by a neuron *i*, N_i^{level} is a set of pre-synaptic neurons that project onto neuron *j* on level (*level*), and ρ_{ji}

denotes the plasticity of the link between pre-synaptic neuron i and post-synaptic neuron j, as shown in Fig. 3(a).

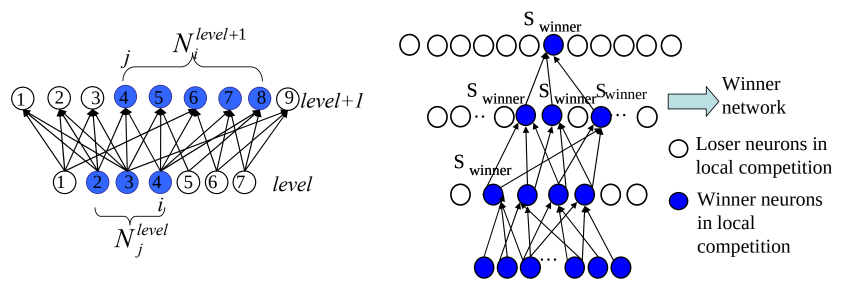

Fig. 3. (a) Interconnection structure to determine a local winner, (b) The winner network

Such local competition can be easily implemented using a current-mode WTA circuit [12]. A local winner neuron, for example $N_4^{level+1}$ in Fig. 3(a), will pass its signal strength to its pre-synaptic neuron N_4^{level}, and all other post-synaptic branches connecting neuron N_4^{level} with the losing nodes will be logically cut off. Such local competition is done first on the highest level. The signal strengths of neurons which win in their corresponding local competitions propagate down to the lower levels and the same procedure continues until the first input layer is reached.

The global winning neuron on the top level depends on the results of all local competitions. Subsequently, the signal strength of the global winner is propagated down to all lower-level neurons which connect to the global winner. Most of the branches not connected to the global winner are logically cut off, while the branches of the global winner are kept active. All the branches that propagate the local winner signal down the hierarchy form the **winner network**, as shown in Fig. 3(b).

Depending on the connectivity structure, one or more winner networks can be found. By properly choosing the connectivity structure, we may guarantee that all of the input neurons are in a single winner network so that the output level contains a single winner.

Let us use a 3-layer winner network (1 input level, 2 secondary levels and 1 output level) as an example. The network has 64 primary input neurons and 4096 output neurons with 256 and 1024 secondary neurons, respectively. The number of active neurons in the top level decreases with increasing numbers of input connections. As shown in Fig.4, when the number of input links to each neuron is more than 8, a single winner neuron in the top level is achieved.

Since the branches logically cut off during local competition will not contribute to post-synaptic neuronal activities, the synaptic strengths are recalculated only for branches in the winner network.

As all the branches of the winner network are used, the signal strength of pathways to the global winner are not reduced. However, due to the logically disconnected branches, the signal strength of pathways to other output neurons are suppressed. As a result, an input pattern activates only some of the neurons in the winner networks. The weights are only adjusted using Hebbian learning for links in winner networks to reinforce the activation level of the global winner. After updating, weights are scaled so that they are still spread on the unit-sphere.

In general, the winner network with secondary neurons and sparse connections, builds sparse representations in three steps: sending data up through the hierarchy, finding the winner network and global winner by using local competitions, and training. The winner network finds the global winner efficiently without iterations usually adopted in MAXNET [8][9]. It provides an effective and efficient solution to the problem of finding global winners in large networks. The advantages of sparse winner networks are significant for large size memories.

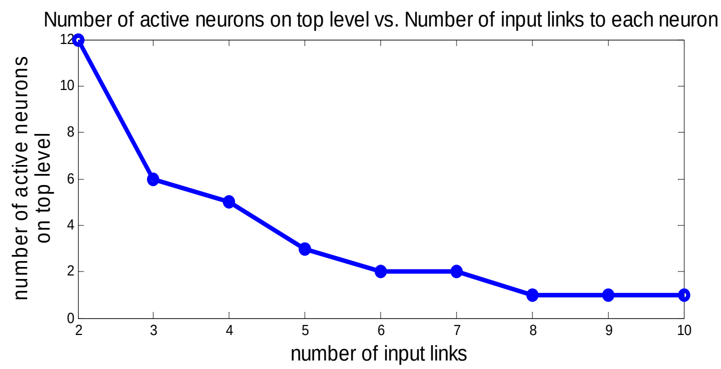

Fig. 4. Effect of number of input connections to neurons

3 Winner Network with Oligarchy-Takes-All

The recognition using a single-neuron representation scheme in the winner network can easily fail because of noise, fault, variant views of the same object, or learning of other input patterns due to an overlap between activation pathways. In order to have distributed, redundant data representations, an OTA network is proposed in this work to use a small group of neurons as input representations.

In an OTA network, the winning neurons in the oligarchy are found directly in a feed-forward process instead of the 3-step procedure used in the winner network as described in section 2. Neurons in the 2^{nd} layer combine weighted inputs and use a threshold activation function as in the winner network. Each neuron in the 2^{nd} layer competes in a local competition. The projections onto losing nodes are logically cut off. The same Hebbian learning as is used in the winner network is carried out on the logically connected links. Afterwards, the signal strengths of the 2^{nd} level are recalculated considering only effects of the active links. The procedure is continued until the top level of hierarchy is reached. Only active neurons on each level are able to send the information up the hierarchy. The group of active neurons on the top level provides redundant distributed coding of the input pattern. When similar patterns are presented, it is expected that similar groups of neurons will be activated. Similar input patterns can be recognized from the similarities of their highest level representations.

4 Experimental results

The learning abilities of the proposed models were tested on the 3-layer network described in section 2. The weights of connections were randomly initialized within the range [-1, 1]. A set of handwritten digits from the benchmark database [13] containing data in the range [-1, 1] was used to train the winner network or OTA networks. All patterns have 8 by 8 grey pixel inputs, as shown in Fig. 5. Each input pattern activates between 26 and 34 out of 4096 neurons on the top level. The groups of active neurons in the OTA network for each digit are shown in Table 1. On average, each pattern activates 28.3 out of 4096 neurons on the top level with the minimum number of 26 neurons and the maximum number of 34 neurons.

Fig. 5. Ten typical patterns for each digit

Table 1. Active neuron index in the OTA network for handwritten digit patterns

digit	Active Neuron index in OTA network							
0	72	91	365	371	1103	1198	1432	1639	...
1	237	291	377	730	887	1085	1193	1218	...
2	294	329	339	771	845	1163	1325	1382	...
3	109	122	237	350	353	564	690	758	...
4	188	199	219	276	307	535	800	1068	...
5	103	175	390	450	535	602	695	1008	...
6	68	282	350	369	423	523	538	798	...
7	237	761	784	1060	1193	1218	1402	1479	...
8	35	71	695	801	876	1028	1198	1206	...
9	184	235	237	271	277	329	759	812	...

The ability of the network to classify was tested by changing 5 randomly selected bits of each training pattern. Comparing the OTA neurons obtained during training with those activated by the variant patterns, we find that the OTA network successfully recognizes 100% of the variant patterns. It is expected that changing more bits of the original patterns will degrade recognition performance. However, the tolerance of the OTA network for such change is expected to be better than that of the winner network. Fig. 6 compares the performances of the winner network and the OTA network for different numbers of changed bits in the training patterns based on 10 Monte-Carlo trials. We note that increasing the number of changed bits in the patterns quickly degrades the winner network's performance on this recognition task. When the number of bits changed is larger than 20, the recognition correctness stays around 10%. However,10% is the accuracy level for random recognition for 10 digit patterns recognition. It means that when the number of changed bits is over 20, the winner network is not able to make useful recognition. As anticipated, the OTA network has much better fault tolerance and it is resistant to this degradation of recognition correctness.

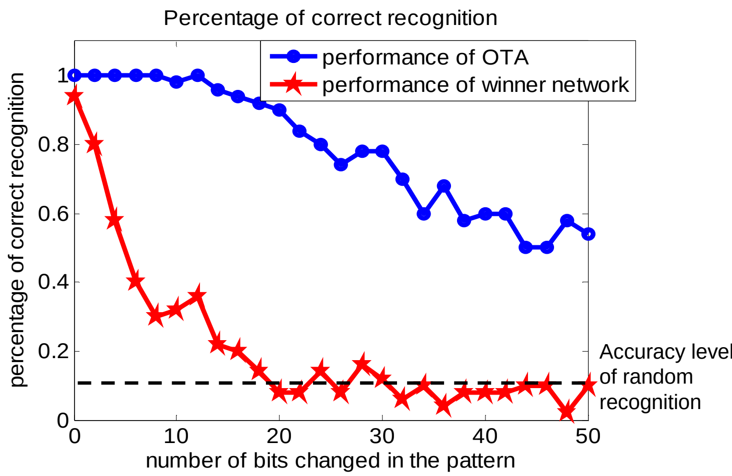

Fig. 6. Recognition performance of the OTA network and the winner network

5. Conclusions

This paper investigates a mechanism for reliably producing sparse coding in sparsely connected networks and building high capacity memory with redundant coding into sensory pathways. Activity reduction is accomplished with local rather than global competition, which reduces hardware requirements and computational cost of self-organizing learning. High memory capacity is obtained by means of layers of secondary neurons with optimized numbers of interconnections. In the winner network, each pattern activates a dominant neuron as its representation. In the OTA network, a pattern triggers a distributed group of neurons. With OTA, information is redundantly coded so that recognition is more reliable and robust. The learning ability of the winner network is demonstrated using experimental results. The proposed models produce features of a learning memory that may prove essential for developing machine intelligence.

Reference:

1. Starzyk, J.A., Liu, Y., He, H., Challenges of Embodied Intelligence, Proc. Int. Conf. on Signals and Electronic Systems, ICSES'06, Lodz, Poland, Sep. 17-20 (2006)
2. Kandel, E. R., Schwartz, J. H., Jessell, T. M., Principles of Neural Science, McGraw-Hill Medical; 4th edition (2000)
3. Anderson, J., Learning in Sparsely Connected and Sparsely Coded System, Ersatz Brain Project working note, (2005)

4. Barlow, H. B., Single Units and Sensation: A Neuron Doctrine for Perceptual Psychology? Perception, Vol.1, (1972) 371-394
5. Amari, S, Neural Representation of Information by Sparse Encoding, Brain Mechanisms of Perception and Memory from Neuron to Behavior, Oxford University Press, (1993) 630-637
6. Földiak, P., Young, M. P., Sparse Coding in the Primate Cortex, The Handbook of Brain Theory and Neural Networks, The MIT Press, (1995) 895-898
7. Olshausen, B. A., Field, D. J., Sparse coding of sensor inputs, Current Opinions in Neurobiology, Vol.14, (2004) 481-487
8. Haykin, S., Neural Networks: A Comprehensive Foundation, Prentice Hall (1999)
9. Zurada, J. M., Introduction to Artificial Neural Systems, West Publishing Company (1992)
10. Vogel, D. D., Boos, W., Sparsely connected, Hebbian networks with strikingly large storage capacities, Neural Networks, Vol. 10, No. 4 (1997) 671-682
11. D. D. Vogel, A neural network model of memory and higher cognitive functions, Int J Psychophysiol, Vol. 55 (1) (2005) 3-21
12. Starzyk, J. A. Fang, X., A CMOS Current Mode Winner-Take-All Circuit with both Excitatory and Inhibitory Feedback, Electronics Letters, Vol.29 (10) (1993) 908-910
13. LeCun, Y., Cortes, C., The MNIST database of handwritten digits, http://yann.lecun.com/exdb/mnist/