Incremental Self-Organizing Map (iSOM) in Categorization of Visual Objects

Andrew P. Papliński

Monash University, Australia
Andrew.Paplinski@monash.edu

Abstract. We present a modification of the well-known Self-Organizing Map (SOM) in which we incrementally allocate the neuronal nodes to progressively added new stimuli. Our incremental SOM (iSOM) aims at the situation when a stimulus, or percept, is represented by a number of neuronal nodes a typical case in biological situation when the redundancy of representation of data is important. The iSOM is applied to categorization of visual objects using the recently introduced feature vector based on the angular integral of the Radon transform [10].

Keywords: Self-organizing maps, Incremental learning, Radon transform.

1 Introduction

We present a modification of the well-known SOM in which we incrementally allocate the neuronal nodes to new stimuli, or data points. The incremental SOM is applied to categorization of visual objects using the feature vector based on the angular integral of the Radon transform.

Self-organizing maps [7] have found numerous applications due to the fact that they are able to perform non-linear dimensionality reduction [8,14] mapping high-dimensional data (observation points, stimuli) $\mathbf{x} \in \mathbb{R}^D$ into a low-dimensional (typically 2) "position" vector $\mathbf{v} \in \mathbb{R}^2$ in the latent space, so that

$$\mathbf{v} = \mathcal{K}(W \cdot \mathbf{x}) \tag{1}$$

where W is an $M \times D$ matrix of parameters, the weight matrix, M being the number of nodes (neurons), and $\mathcal{K}$ is the Winner-Takes-All function identifying the position of the neuronal node $\mathbf{v}$ for which the post-synaptic activity $W \cdot \mathbf{x}$ attains the maximum. In this formulation, each k-th neuronal node, $k \in \{1 \dots M\}$, is identified by its weight vector $\mathbf{w}_k$ and the position vector $\mathbf{v}_k$. By applying a learning law which successfully drags weights vectors $\mathbf{w}$ towards data points $\mathbf{x}$, the mapping (1) can achieve the topology preserving property, mapping close data points $\mathbf{x}$ to the common node vector $\mathbf{v}$, hence, performing the clustering.

We start with a brief review of fundamental concepts related to typical applications of SOMs and their modifications, to locate the place where our iSOM fits in the large family of self-organizing algorithms. In the simplest case, the

T. Huang et al. (Eds.): ICONIP 2012, Part II, LNCS 7664, pp. 125–132, 2012.

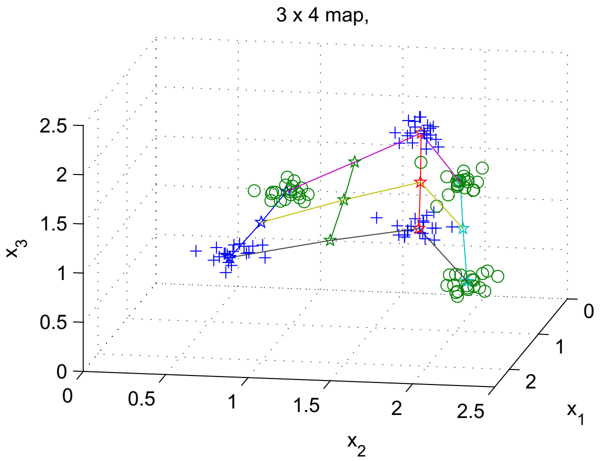


Fig. 1. Clustering with a SOM. A SOM with 3×4 nodes approximates 120 3-D data points arranged in six clusters.

number of data points, N, and the number of the neuronal nodes, M, are fixed. Typically, when clustering applications are involved, the number of neuronal nodes is smaller than the number of the data points and the SOM "explains the data" giving a simplified, 2-dimensional view of the D-dimensional data as in Figure 1. Note that some of the nodes in Figure 1 do not really fit any cluster and can be considered as "dead" ones. Since in this example the dimensionality of data is low ($D = 3$), visualization can be performed in the input space.

In categorization applications as in Figure 2, when the dimensionality of the feature vector is high, the mapping is visualised in the latent space when categories are allocated to neuronal nodes. As in the clustering case, if the structure and the size of the SOM is fixed there might be "dead" nodes not representing any particular category. On the other hand, we can see that the category of cats is crowded into a single neuronal node.

The above two generic cases[1] show the fundamental problem related to the rigid structure of the nodes in the latent space, namely, inefficient use of the nodes to represent data. Conceptually, a regular lattice from the latent space is well-suited for the data which is evenly distributed in the input space, a really trivial case. In general, we would like to have the structure and the number of nodes to be adjustable in order to represent better the important aspects of the data.

To address the above problem a number of adjustable SOM structures have been developed. The first, to my knowledge, such a structure is known as the Incremental Grid Growing Feature Map (IGGFM) [2], followed by the Growing Cell Structures (GCS) [4], the Growing Neural Gas (GNG) [5], the Dynamic

[1] An interesting case, when the number of nodes is equal to the number of data points (image pixels) is shown on our web page `http://www.csse.monash.edu.au/~app/`.

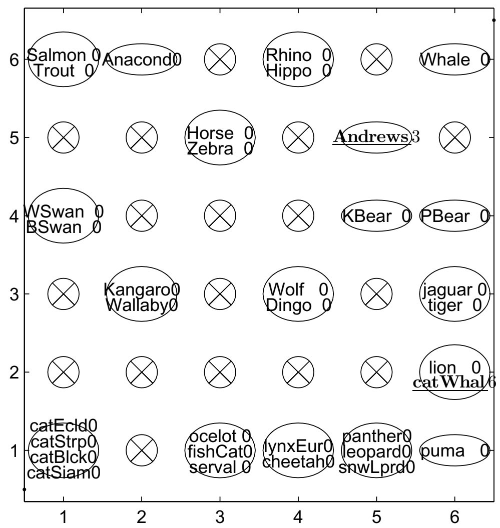


Fig. 2. Categorization with a SOM. A SOM with 6×6 nodes is used to categorized 32 "objects" (animals). Two test animals are also shown. The numbers are the relative distance from the winning node. The "dead" nodes are marked with crosses.

aka Growing SOM (GSOM) [1], and the Time-Adaptive SOM (TASOM) [16]. In the above self-organizing algorithms the nodes are located on a regular lattice ([2,1,16]), or on a lattice built from the triangles ([4,5]) with the objective for the network to map the topology, or probability distribution of the data. Extension of the above concepts to the multi-map network is presented in [17]. Here the reader can find a good description of the above concepts.

In our recent applications a network of SOMs is used to model multimodal integration [12,6,11] and rotation invariant categorization of visual objects [10]. In the above applications related to modelling selected aspects of cortical functions, we argue that a number of neuronal units is required to represent a concept, or percept, in order to create redundancy, hence ensuring robust working of the cortex. We selected this number to be in the range $\epsilon \in [16 \ldots 20]$ nodes per data unit. The motivation was to have the number of nodes of the same order of magnitude as the number of neurons in the cortical mini-columns. Since the number of data points is smaller than the number of nodes, representation of data topology plays a secondary issue, unlike in the works mentioned above in which it is the primary aspect.

In this paper we consider SOMs in which we progressively add new objects, or data points, and generate a number of new nodes to maintain a fixed density of nodes per object ϵ. Our latent, or neuronal space consists of randomly distributed nodes inside of a unity circle. In other words, with each new object added, we also incrementally generate ϵ new nodes. Hence, we call our extension of Self-Organizing Maps an Incremental SOM, or iSOM.

2 Categorization of Visual Objects

Categorization of visual objects remains one of the grand challenges despite of an enormous progress in this area. The principal problem is the selection of the features that can represent a visual object in a way invariant to factors like rotation, scaling, changes in illumination and the viewpoint. Currently the local descriptors, specifically, based on the Scale Invariant Feature Transform (SIFT) [9], are most commonly used [15,3].

In this paper categorization of the visual objects is only used as an illustration for our incremental SOM. Therefore, we use a simplified feature vector based on the Radon transform as presented in [10]. The Radon transform $Rf(\theta, s)$ [13] of a function $f(z)$, where $z = x + jy$, is calculated as an integral of $f(z)$ over straight lines $z(t) = e^{j\theta}(s + jt)$ with the slope θ

$$Rf(\theta, s) = \int_{-\infty}^{+\infty} f(e^{j\theta}(s + jt))dt \tag{2}$$

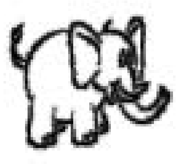
where s is the distance of the line from the origin and t is its parameter. In order to obtain a feature vector $h(s)$ that is rotation invariant, we integrate transform (2) over all angles

$$h(s) = \int_{0}^{2\pi} Rf(\theta, s)d\theta \tag{3}$$

Such a feature (signature) function $h(s)$ retains some characteristics of the original image $f(x, y)$, but the angular dependency is removed, hence the rotational invariance is achieved.

In the discrete case, the feature vector based on eqn (3) has the size equal to the diagonal of the image, which is significantly smaller than the size of the SIFT-based feature vectors. In what follows we use two types of images:

– The characters represented as binary 46×46 images generated using the Times New Roman font of size 24. There are 26 lower-case characters. The feature vector calculated using eqn (3) has the size of 67.
– The 64×64 gray scale iconic images of 30 animals as in the example shown on the right. The feature vector is 82-element long.



The feature vectors are normalised to be in the range $[-1, +1]$ and then projected up on the $(D + 1)$-dimensional hypersphere, so that we operate with the unity vectors.

3 Incremental SOM

As it has been said above our iSOM aims at situation where the number of nodes per stimulus, ϵ is approximately constant and in the range $[16 \ldots 20]$. The nodes are randomly distributed on a top half-surface of a unity sphere. In Figure 3 the

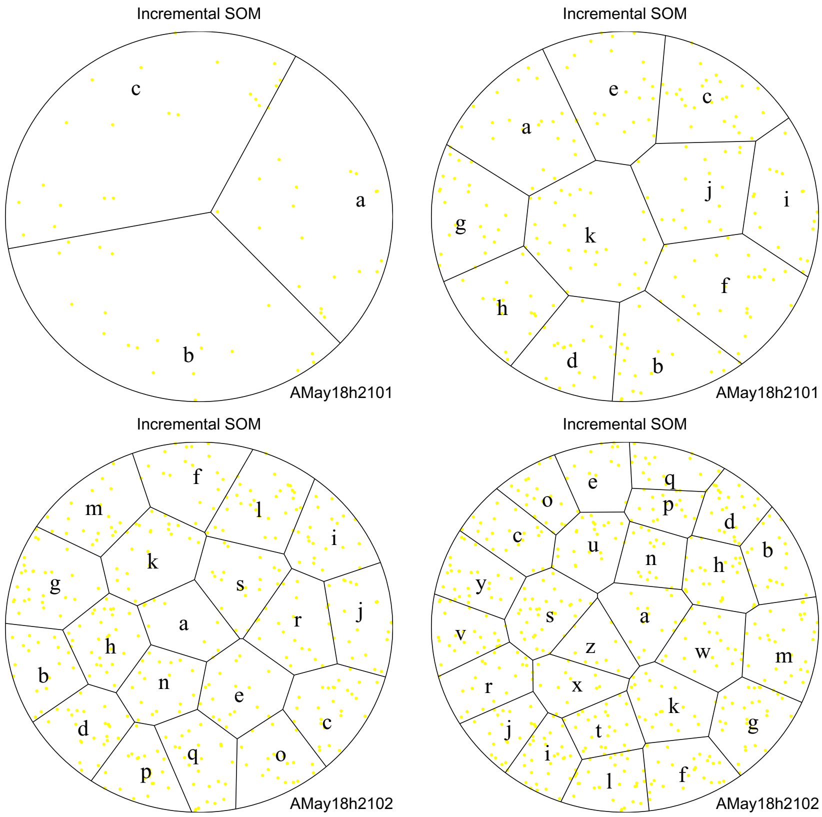


Fig. 3. The evolution of the map after application of 3, 11, 19 and 26 visual stimuli (letters)

location of the nodes projected onto a unity circle are marked with (yellow) dots. Each neuronal node is identified by its weight vector $\mathbf{w}_k$ and the position vector $\mathbf{v}_k$, both being unity vectors. We start with some initial number of stimuli (three in our example) and nodes $(3 * \epsilon)$. The weight vectors are intialised around the north pole of the hypersphere, and a standard Kohonen "dot-product" learning law is applied in which the update of the weight vector $\mathbf{w}_j$ for the jth neuron is described by the following expression:

$$\Delta\mathbf{w}_j = \eta \cdot \Lambda_j \cdot (\mathbf{x}^T - d_j \cdot \mathbf{w}_j) \; ; \quad d_j = \mathbf{w}_j \cdot \mathbf{x} \tag{4}$$

where Λ_j is a neighbourhood function, Gaussian in our case, centred on the position of the winning neuron, and d_j is the **post-synaptic activity** of the jth neuron. Similarly to the neighbourhood function, the gain η is also reduced according to a Gaussian curve, $\eta = \exp(-n^2/(2\sigma))$, where σ is selected so that $\eta = 0.5$ for $n = E/2$, E being the number of epochs. This ensures a good proportion between the ordering and convergence phases.

After learning is completed for a given number of stimuli, say 11, (see Figure 3), we choose a required number of new stimuli, say γ, one in particular, and generate $\gamma \times \epsilon$ new nodes as described above. For simplicity, the new weight vectors are intialised as before around the north pole, and the learning procedure is repeated. Alternatively, the weight vectors of the newly generated neuronal nodes could be cloned from the nearest neighbours. The results of learning are shown in Figure 3. As expected, at each stage the map organizes the stimuli according to their visual features, e.g., keeping 'f', 'l', and 'i' together.

The second example is similar and we incrementally create map for categorization of animals base on their visual appearances. The resulting maps after application of 3, 12, 21 and 30 stimuli are shown in Figure 4. Unlike in the letter case, this time we categorize visual objects represented by the gray scale images. It is more difficult to notice similarities with respect to the feature vector as in eqn (3). Since the images were not standardized with respect of the average value of the gray scale pixels, the lighter images of objects tend to be clustered together, as do the darker images.

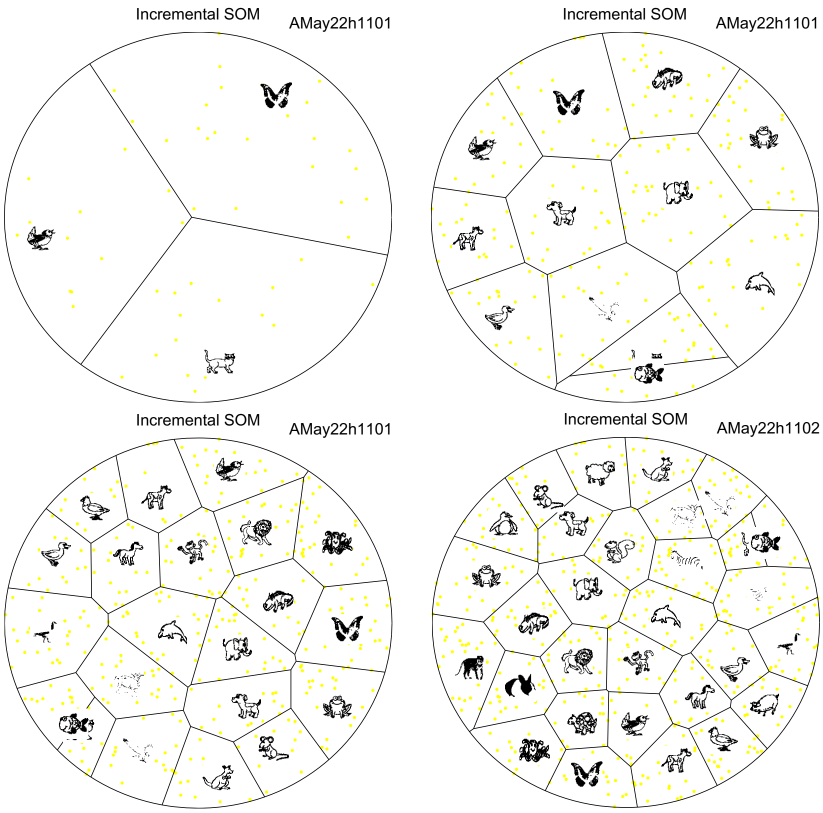


Fig. 4. The evolution of the map after application of 3, 12, 21 and 30 visual stimuli (images of animals)

4 Concluding Remarks

We have presented a new version of a self-organizing map, incremental SOM, targeted at the situation when percepts represented by stimuli, or data points, are mapped onto a group of neuronal modes thus ensuring redundancy required in the biological systems. The mapping is performed incrementally, so that, for each new stimulus applied to the map, a statistically fixed number of nodes is randomly generated in the latent space. We apply the new iSOM to categorization of visual objects testing a newly introduced feature vector based on the angular integral of the Radon transform of the image representing the visual object.

Finally, with reference to Figures 3 and 4, we would like to emphasize that the "position" vector $\mathbf{v}$ describes the location of the node in the latent space, not in the physical space. It means that the physical location of neurons is not affected by the incremental learning process. The "position" vector $\mathbf{v}$, similarly to the weight vector $\mathbf{w}$, is a property of a neuron, not its physical location. In this sense, the neurons are not "crowded" when the new percepts are introduced.

References

1. Alahakoon, D., Halgamuge, S.K., Srinivasan, B.: Dynamic Self-organizing Maps with Controlled Growth for Knowledge Discovery. IEEE Trans. Neural Networks 11(3), 601–614 (2000)
2. Blackmore, J., Miikkulainen, R.: Incremental Grid Growing: Encoding High-Dimensional Structure into a Two-Dimensional Feature Map. In: Proc. IEEE Int. Conf. Neural Networks, pp. 450–455 (1993)
3. Bosch, A., Zisserman, A., Munoz, X.: Image Classification Using Random Forests and Ferns. In: 11th Int. Conf. Computer Vision, vol. 21(4), pp. 1–8 (2007)
4. Fritzke, B.: Growing Cell Structures — a Self-Organizing Network for Unsupervised and Supervised Learning. Neural Networks 7(9), 1441–1460 (1994)
5. Fritzke, B.: A Growing Natural Gas Network Learns Topologies. In: Advances in Neural Information Processing, vol. 7, pp. 625–632. MIT Press (1995)
6. Jantvik, T., Gustafsson, L., Popliński, A.P.: A Self-Organized Artificial Neural Network Architecture for Sensory Integration with Applications to Letter–Phoneme Integration. Neural Computation 23, 2101–2139 (2011)
7. Kohonen, T.: Self-Organising Maps, 3rd edn. Springer, Berlin (2001)
8. Lee, J.A., Verleysen, M.: Nonlinear Dimensionality Reduction. Springer (2007)
9. Lowe, D.G.: Distinctive Image Features From Scale-Invariant Keypoints. Int. J. Comp. Vision 60(2), 91–110 (2004)
10. Papliński, A.P.: Rotation Invariant Categorization of Colour Images Using Radon Transform. In: Proc. WCCI–IJCNN, pp. 1408–1413. IEEE (2012)
11. Papliński, A.P., Gustafsson, L., Mount, W.M.: A Model of Binding Concepts to Spoken Names. Aust. Journal of Intelligent Information Processing Systems 11(2), 1–5 (2010)
12. Papliński, A.P., Gustafsson, L., Mount, W.M.: A Recurrent Multimodal Network for Binding Written Words and Sensory-Based Semantics into Concepts. In: Lu, B.-L., Zhang, L., Kwok, J. (eds.) ICONIP 2011, Part I. LNCS, vol. 7062, pp. 413–422. Springer, Heidelberg (2011)

13. Radon, J.: Radon transform, http://en.wikipedia.org/wiki/Radon_transform
14. Sammon, J.W.: A Nonlinear Mapping for Data Structure Analysis. IEEE Trans. Computers 18(5), 401–409 (1969)
15. Serre, T., Wolf, L., Poggio, T.: Object Recognition with Features Inspired by Visual Cortex. In: IEEE Comp. Soc. Conf. Comp. Vision and Patt. Recognition, vol. 2, pp. 994–1000 (2005)
16. Shah-Hosseini, H., Safabakhsh, R.: TASOM: A New Time Adaptive Self-Organizing Map. IEEE Trans. Syst., Man, Cyber. B 33(2), 271–282 (2003)
17. Shah-Hosseini, H., Safabakhsh, R.: Binary Tree Time Adaptive Self-Organizing Map. Neurocomputing 74, 1823–1839 (2011)