Self-Organizing Map Formation: Foundations of Neural Computation

Edited by Klaus Obermayer and Terrence J. Sejnowski

A Bradford Book

The MIT Press
Cambridge, Massachusetts
London, England

Introduction

The papers on self-organizing map formation collected here have appeared in *Neural Computation* over the past ten years. The papers provide an overview of the field as well as recent developments and can be used by students, researchers, and practitioners. The field of self-organizing maps is a branch of unsupervised learning, whose goal is to determine the statistical properties of input data without explicit feedback from a teacher. A previous volume in this series (Hinton and Sejnowski 1999) collected additional papers from *Neural Computation* on unsupervised learning that are not concerned with self-organizing map formation.

Since much of the original inspiration for the self-organizing maps arose from biological studies, this volume begins with a series of papers that attempts to model the organization of cortical maps and ends with a series of papers on the theory and the applications of the related artificial neural network algorithms. These papers illustrate the fruitfulness of interactions between the modeling of biological systems and the exploration of practical algorithms for solving difficult computational problems. Models of biological systems provide inspiration and a source of new ideas for unsupervised data analysis; the theoretical analyses of the derived algorithms in turn give rise to new hypotheses for the development and function of cortical maps.

Topographic Maps in the Cerebral Cortex

A central organizing principle in the mammalian cerebral cortex is the orderly topographical arrangement of sensory and motor neurons with similar response properties across the cortical surface. Anatomical evidence for cortical maps, based on lesion studies and electrical stimulation, has been available since the early nineteenth century. When methods for recording from single cortical neurons were developed in the 1960s, it became possible to determine how neurons were organized in the cortex. Vernon Mountcastle was the first to report that in a vertical column all neurons tend to have the same properties, based on recording from neurons in the somatosensory cortex. In the visual cortex, David Hubel and Torsten Wiesel reported smooth changes in the response properties of cortical neurons, such as the ocular dominance and orientation of an edge (Hubel and Wiesel 1962). Since then cortical maps have been found and characterized in nearly all of the sensory and motor areas of the brain (Churchland and Sejnowski 1992).

Cortical maps, particularly in the visual cortex, have been the focus of studies into activity-driven mechanisms for synaptic plasticity and their interplay with non-activity-dependent processes based on cell adhesion molecules and diffusible substances (Goodman and Shatz 1993; Quartz and Sejnowski 1997). Recently, new imaging methods (Blasdel and Salama 1986; Grinvald et al. 1986) have made it possible to monitor the distribution of neuronal selectivities across the cortex and over time. These methods have allowed the large-scale development of neuronal response properties to be studied with much greater coverage than previously possible with single-unit recording techniques. Driven by these experimental results, a series of computational studies have attempted to explain the structure of the cortical maps by simulating and analyzing cortical models. One set of studies was concerned with the mechanisms underlying development and plasticity (see chapter 8 and Swindale 1996 for reviews). These approaches were typically based on constrained Hebbian learning and intracortical competition; despite their simplicity, they were able to explain many features of the experimentally observed spatial patterns.

Another set of studies was concerned with the computational advantages of these maps. For example, "minimal wiring" (Durbin and Mitchison 1990), which seeks to reduce the volume of axons needed, has been used to explain the orderly architecture of the cerebral cortex and the patterns of discontinuities (primarily the visual cortex); analogies with parallel computers (Nelson and Bower 1990) have suggested an interpretation of cortical maps in terms local processing, load balancing, and communication requirements. Other approaches based on the goal of efficient coding (Barlow 1961; Field 1994) have attempted to explain the properties of receptive fields (Atick 1992) and their spatial arrangement (Ritter et al. 1991) in terms of the statistics of the signals and the noise. In contrast to the developmental studies, however, most of these modeling studies were less successful in explaining the spatial layout of the maps, and there is still no adequate explanation for why the cortex has such a regular structure.

Self-Organizing Maps

Inspired by the ubiquity of cortical maps in the central nervous system, several mapping algorithms were introduced in the 1970s and early 1980s (Takeuchi and Amari 1979; Whitelaw and Cowan 1981; Grossberg 1976; Malsburg 1973; Kohonen 1982b; Kohonen 1982a; Willshaw and von der Malsburg 1976). It soon became apparent that self-organizing maps are valuable tools for unsupervised data analysis (Kohonen 1987; Ritter, Martinetz, and Schulten 1992). Subsequently, self-organizing maps were applied to real-world problems such as preprocessing for solving classification and regression problems or as a technique for extracting and visualizing salient features in the data (Kohonen 1995; Kohonen 1997).

Although mapping algorithms became widely adopted in the neural network community because of their conceptual simplicity and their computational efficiency, theoretical analysis lagged. Important advances have recently been made by formulating mapping algorithms in terms of cost-functions (Cottrell and Fort 1987; chapter 13; chapter 5), which allow a deeper theoretical understanding and suggest improved optimization procedures. The introduction of cost functions also made it possible to link mapping techniques to methods in classical statistics, which opened new approaches through generative models as well as through statistical techniques such as maximum likelihood and Bayesian analysis (chapter 15).

Receptive Fields

Three papers included here illustrate three aspects of receptive field development that are important for the formation of cortical maps: Hebbian learning, efficient encoding, and statistical relationships. The analysis of constrained Hebbian learning for the activity-driven formation of receptive fields in the primary visual cortex in MacKay and Miller (chapter 1) had a considerable impact on subsequent analyses of map formation. Li and Atick (chapter 2) examine the computational advantages of these emerging receptive fields and show that the filters found in the early visual system, by reducing second-order correlations in the visual input, form an efficient representation of the natural visual environment, which is to some extent invariant under translation and scale. Yuille, Smirnakis and Xu (chapter 3) focus on the emergence of feature detectors based on higher-order correlations in the input data. They examine the receptive fields that emerge by minimizing the Kullback-Leibler divergence between the true probability distribution of a signal amid distracting noise and the output probability distribution of the neurons in the network, thus forming a link with Bayesian theories of visual perception.

Models of Topographic Maps in the Brain

Another group of papers is devoted to an analysis of topographic maps and their formation via objective functions. Zhang (chapter 4) analyzes and extends a continuum model (Amari 1983) and derives point-spread resolution, magnification factors, and bandwidth resolution for topographic maps. Map formation can be characterized by a Lyapunov function that is minimized during map formation. Goodhill and Sejnowski (chapter 5) relate objective functions used by several authors in models of topographic map formation and show that they can be seen as particular cases of a more general cost function that they call the C-measure. Wiskott and Sejnowski (chapter 6) compare objective functions for map formation that use Hebbian learning and relate different models through coordinate transformations. This chapter also discusses additional constraints imposed on the connection strengths and shows how they should be chosen to be consistent with the objective function. Linsker (chapter 7) derives topographic maps from an information-theoretic argument based on "efficiency": maximizing the mutual information between the network input and output. This infomax approach has influenced other unsupervised learning algorithms for extracting higher-order information from the input data (Bell and Sejnowski 1995; Lee, Girolami, and Sejnowski 1999).

Cortical Maps of Stimulus Features

Among the most prominent architectural elements of primary visual cortex are orientation selectivity, ocular dominance, and (in primates) color blobs. The mapping of orientation selectivity and ocular dominance have received most of the attention, although other features such as disparity have also been examined (Berns, Dayan, and Sejnowski 1993). Erwin, Obermayer, and Schulten (chapter 8) provide a systematic review of developmental models for orientation selectivity and ocular dominance and compare model predictions with experimental data from macaque striate cortex. Bauer (chapter 9) considers the influence of areal boundaries on the global map structure for the example of ocular dominance bands, and Piepenbrock, Ritter and Obermayer (chapter 10) revisit constraints in models of Hebbian learning to show that these nonlinearities are essential for the joint development of orientation selectivity and ocular dominance maps. Barrow, Bray, and Bud (chapter 11) apply Hebbian learning to a three-layer recurrent network trained with patches from color photographs of natural scenes and compare the results to a principal component analysis of the image data. Not only did their network develop spatial filters similar to the principal components of the data, but it also grouped the corresponding feature-detecting cells within the cortical layer into color "blobs" embedded in a sea of contrast-sensitive, orientation-selective cells. The resulting pattern is similar to what is seen in the superficial layers of visual cortex area V1 in primates. Mitchison (chapter 12) makes the assumption that local cortical processing occurs within feature spaces and explores the arrangement of computational elements across cortex such that the "wiring length" for connections between them is minimal. The results mathematically relate the elastic net to the self-organizing map approaches.

Self-Organizing Maps for Unsupervised Data Analysis

In a seminal paper, Luttrell (chapter 13) used Bayesian methods to relate self-organizing map formation to the training of probabilistic autoencoders. This insight led to the formulation of a cost function from which Kohonen's original self-organizing map approach could be derived as an approximation of the expectation-maximization (EM) optimization of this cost function; it also paved the way for several extensions of the self-organizing map. Utsugi (chapter 14) formulated a generalized deformable model via a Gaussian mixture model and derived a "mapping" algorithm by a maximum a posteriori estimate of the weight parameters. This is the first paper to apply Bayesian model selection techniques to estimate the hyperparameters for the noise and the prior on the weights. Bishop, Svensen, and Williams (chapter 15) follow up on the idea of generative models. They consider a small number of unobservable, "explanatory" variables that generate the observed distribution of the data, and use statistical techniques to infer their values for each data point, thus performing dimension reduction. Mulier and Cherkassky (chapter 16) relate self-organizing maps to classical nonparametric regression.

Extensions of Self-Organizing Maps

Graepel and Obermayer (chapter 17), following up on Luttrell (chapter 13) and Hofmann and Buhmann (1997), consider dissimilarity values between data items and apply a self-organizing map to combine clustering with metric multidimensional scaling. Kohonen, Kaski, and Lappalainen (chapter 18) combine the mapping property of the self-organizing map with a local projection method, which is similar to principal component analysis, and apply this method to temporal sequences of data. Lin, Grier, and Cowan (chapter 19) successfully use coupled one-dimensional self-organizing maps for the blind separation of sources, while Bruske and Sommer (chapter 20) apply growing and pruning algorithms to self-organizing maps. This section concludes with two surprising applications of mapping algorithms to standard computer science problems: combinatorial optimization and sorting. Durbin, Szeliski, and Yuille (chapter 21) use an elastic net method to find solutions for the traveling salesman problem, and Budinich (chapter 22) shows how self-organizing maps can beat quicksort.

Conclusions

Self-organizing maps occur in both biological and engineering domains. The papers in the collection highlight some of the most important issues that arise in studying the properties of these maps. In particular, maps can be constructed from relatively simple rules governing the local interactions between neighboring elements. This may allow a wide range of properties to be encoded in sensory maps. Most of modeling studies have focused on maps in the earliest stages of sensory processing, where topography is the dominant organizing principle. At higher stages of processing, the sizes of receptive fields become larger, the topography becomes less distinct, and the mapping of higher-order features rather than topography become a dominant organizing principle (Tanaka 1996). Feedback connections between these areas allows the highest levels to influence the earliest levels of processing. Modeling these extrastriate areas and the interactions between them provides a challenge for future studies that build on the papers collected in this volume. A self-organizing multilevel system would have many practical applications.

References

Amari, S. 1983. Field theory of self-organizing neural nets. *IEEE Trans. SMC* **13**, 741–748.

Atick J. J. 1992. Could information theory provide an ecological theory of sensory processing? *Network* 3, 213–251.

Barlow, H. 1961. The coding of sensory messages. In *Current Problems in Animal Behavior*. Cambridge: Cambridge University Press.

Bell, A. J., and Sejnowski, T. J. 1995. An information-maximization approach to blind separation and blind deconvolution. *Neural. Comp.* 7, 1129–1159.

Berns, G. S., Dayan, P., and Sejnowski, T. J. 1993. A correlational model for the development of disparity selectivity in visual cortex that depends on prenatal and postnatal phases. *Proceedings of the National Academy of Sciences U.S.A.* **90**, 8277–8281.

Blasdel, G. G., and Salama, G. 1986. Voltage sensitive dyes reveal a modular organization in monkey striate cortex. *Nature* **321**, 579–585.

Churchland, P. S., and Sejnowski, T. J. 1992. *The Computational Brain*. Cambridge, MA: MIT Press.

Cottrell, M., and Fort, J. C. 1987. Etude d'un processus d'auto-organisation. *Ann. Inst. Henri Poincaré* 23, 1–20.

Durbin, R. and Mitchison, G. 1990. A dimension reduction framework for understanding cortical maps. *Nature* **343**, 644–647.

Field, D. 1994. What is the goal of sensory coding? *Neural Computation* **6**, 559–601.

Goodman, C., and Shatz, C. 1993. Developmental mechanisms that generate precise patterns of neuronal connectivity. *Cell* 72, 77–89.

Grinvald, A., Lieke, E., Frostig, R. D., Gilbert, C. D., and Wiesel, T. N. 1986. Functional architecture of cortex revealed by optical imaging technique. *Nature* **324**, 361–364.

Grossberg, S. 1976. Adaptive pattern classification and universal recoding, I: Parallel development and coding of neural feature detectors. *Biol. Cybern.* **23**, 121–134.

Hinton, G. E., and Sejnowski, T. J., eds. 1999. *Unsupervised learning: Foundations of neural computation*. Cambridge, MA: MIT Press.

Hofmann, T., and Buhmann, J. 1997. Pairwise data clustering by deterministic annealing. *IEEE PAMI* **19**, 1–14.

Hubel, D., and Wiesel, T. N. 1962. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *J. Physiol. (Lond.)* **160**, 106–154.

Kohonen, T. 1982a. Analysis of a simple self-organizing process. *Biol. Cybern.* **44**, 135–140.

Kohonen, T. 1982b. Self-organized formation of topologically correct feature maps. *Biol. Cybern.* **43**, 59–69.

Kohonen, T. 1987. *Self-organization and associative memory*. New York: Springer-Verlag.

Kohonen, T. 1995. *Self-organizing maps*. Berlin, Heidelberg: Springer-Verlag.

Kohonen, T. 1997. Bibliography on the SOM and the LVQ. Available on the Internet at ⟨http://www.cis.hut.fi/nnrc/refs/⟩.

Lee, T.-W., Girolami, M., and Sejnowski, T. J. 1999. Independent component analysis using an extended infomax algorithm for mixed sub-Gaussian and super-Gaussian Sources. *Neural Computation* **11**(2), 417–441.

Malsburg, C. 1973. Self-organization of orientation sensitive cells in the striate cortex. *Kybernetik* **14**, 85–100.

Nelson, M. E., and Bower, J. M. 1990. Brain maps and parallel computers. *Trends in the Neurosciences* **13**, 403–408.

Quartz, S., and Sejnowski, T. J. 1997. The neural basis of cognitive development: A constructivist manifesto. *Behavioral and Brain Sciences* **20**(4), 537–596.

Ritter, H., Martinetz, T., and Schulten, K. 1992. *Neural computation and self-organizing maps: An introduction*. New York: Addison-Wesley.

Ritter, H., Obermayer, K., Schulten, K., and Rubner, J. 1991. Self-organizing maps and adaptive filters. In E. Domani, J. L. van Hemmen, and K. Schulten, eds., *Physics of Neural Networks*, pp. 281–306. New York: Springer-Verlag.

Swindale, N. 1996. The development of topography in the visual cortex: A review of models. *Network* **7**, 161–247.

Takeuchi, A., and Amari, S. 1979. Formation of topographic maps and columnar microstructures. *Biol. Cybern.* **35**, 63–72.

Whitelaw, V., and Cowan, J. 1981. Specificity and plasticity of retinotectal connections: A computational model. *J. Neurosci.* **1**, 1369–1387.

Willshaw, D. J., and von der Malsburg, C. 1976. How patterned neural connections can be set up by self-organization. *Proc. R. Soc. Lond. B* **194**, 431–445.