Computational models of perceptual learning

Shimon Edelman
School of Cognitive and Computing Sciences
University of Sussex at Brighton
Falmer BN1 9QH, England

Nathan Intrator
Department of Computer Science
Tel-Aviv University
Ramat-Aviv 69978, Israel

March 1999

Abstract

In visual perception, learning is a pervasive phenomenon, which, when properly studied, may offer valuable insights into the inner workings of the brain. We outline a theoretical framework for the computational study of perceptual learning, aiming to make the relationships among the existing models more readily apparent, and to identify promising directions for future research.

1 Introduction

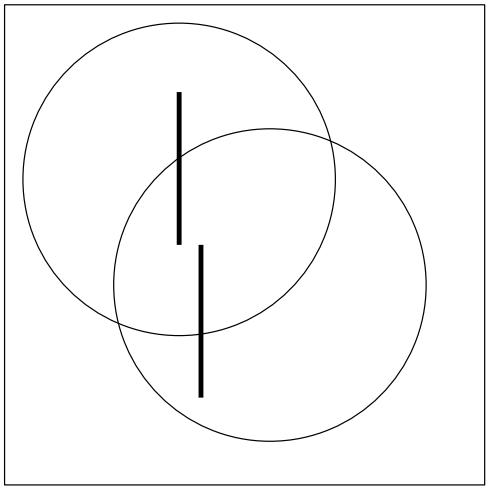
A generation ago, mathematical psychology, which at the time was *the* discipline in charge of modeling behavior, appeared to be in poor shape. William Estes, one of the main *dramatis personae* in that field, described it like this: "Look at our present theories... or at the probabilistic models that are multiplying like overexcited paramecia. Although already too complicated for the average psychologist to handle, these theories are not yet adequate to account for the behavior of a rodent on a runway" (Estes, 1957). During the following decades, when the mainstream psychology underwent a major paradigm shift, the modeling of perceptual learning fared better than what one might have expected from the view expressed by Estes. A new theoretical outlook, which encouraged thinking now termed representational or computational, took over the field. At the same time, the models became, if anything, more complex compared to those of 1957.

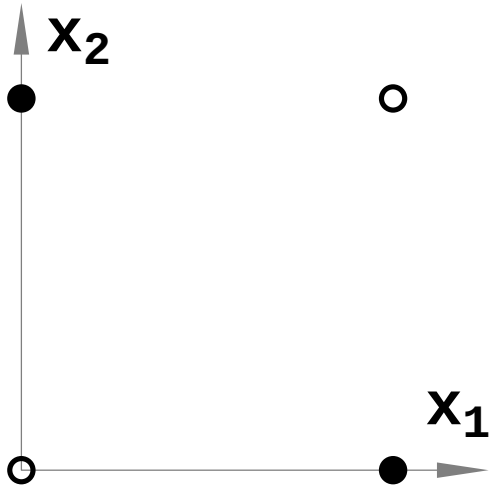
Encouragingly, the models are now also more successful in explaining behavior (rather than merely predicting the probability of a certain response to a given stimulus), while giving no undue troubles to the psychologists.[1] Insofar as there is progress, it seems to stem mainly from (1) the improvement in the experimental techniques that subserve data collection in behavioral and physiological psychology, and (2) the revision of the theoretical basis from which models are drawn. The "rodent on a runway" example mentioned above serves well to illustrate both these points. On the theoretical or conceptual side, the current explanation takes the route presaged by Tolman and based on the concept of cognitive maps (O'Keefe and Nadel, 1978). On the experimental side, the existence of cognitive maps in the rat brain could not have been demonstrated without modern multi-electrode recording methods and the information-processing tools that accompany them.[2]

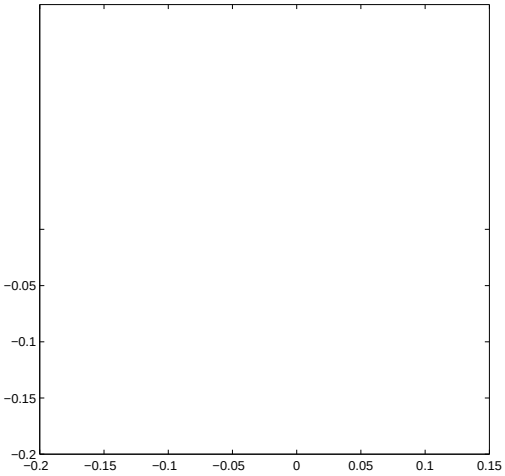
[1] For an interesting historical perspective on these issues, see (Hintzman, 1994).

[2] Cognitive co(tmT18.3(on)-1ide,)TJte

x_1	x_2	y
0	0	0
0	1	1
1	0	1
1	1	0

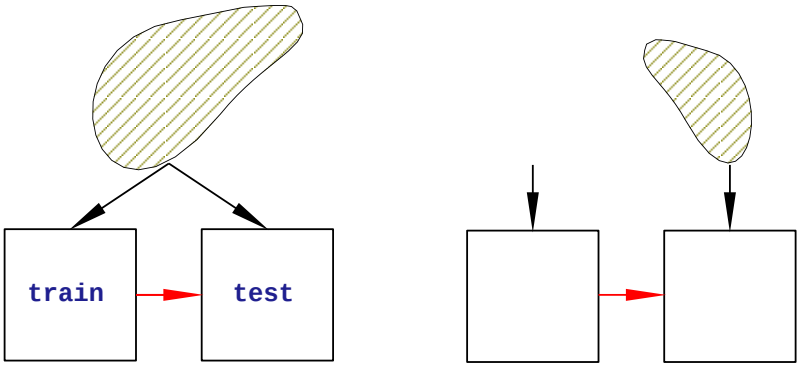






space in such a manner that the probability of generalization between any two stimuli is monotonic in their proximity (i.e., similarity). Shepard's treatment of this issue included a derivation of the monotonic dependence law from some basic assumptions on the probability measure used to quantify generalization.

In theoretical neurobiology, the notion of generalization underlies the "Fundamental Hypothesis" stated in Marr's theory of the cerebral neocortex (Marr, 1970): "Where instances of a particular collection of intrinsic properties (i.e., properties already diagnosed from sensory information) tend to be grouped such that if some are present, most are, then other useful properties are likely to exist which generalize over such instances. Further, properties often are grouped in this way" (pp. 150-151). Although this hypothesis seems at present every bit as convincing as it must have appeared to


train
test

tional capabilities of neural networks. If this foundation is be useful in the development of specific models of learning in the nervous system, statistical samples of stimuli must be shown to contain information necessary for learning. Having been downplayed for decades by the work of Chomsky and his school, the notion that statistical inference can support learning even in markedly "symbolic" domains such as language acquisition is now making a comeback. On the one hand, this process is aided by the growing evidence that humans (both adults and infants) are sensitive to statistical cues present in linguistic stimuli. For example, subjects can extract from such cues, implicitly, information about boundaries between the underlying morphological units (Saffran et al., 1996), word meaning (Markson and Bloom, 1997), and even grammar-like rules (Berns et al., 1997). These findings raise doubts concerning the exclusive applicability of symbol-manipulating computational models to learning problems hitherto coached in purely symbolic terms. Consequently, models built around symbolic abstraction (rule inference) now have to compete routinely with models that posit similarity-based processing (Berry, 1994; Goldstone and Barsalou, 1998).[4]

In visual perception, 3D object recognition is one domain where the hegemony of symbolic/structural models is increasingly challenged by statistical/connectionist learning approaches. Visual recognition gives rise to a variety of learning-related tasks, similar to those encountered in the context of face processing (mentioned briefly above). In these tasks, the lure of symbolic/structural models stems from the observation that for many object classes the main challenge inherent in learning recognition — achieving invariance over transformations or deformations of the stimulus — disappears if the objects are represented structurally (Biederman, 1987). This observation leads to the postulate that learning to recognize an object entails the identification of its parts and the determination of their spatial relationships. Under this assumption, the possession of a library of generic parts that can be assembled in various ways would also endow the system with the ability to represent and process novel objects — the ultimate kind of generalization.

Recently, a different route to invariance and to the ability to process novel shapes has been proposed and implemented in a series of models (Poggio and Edelman, 1990; Edelman and Duvdevani-Bar, 1997; Riesenhuber and Poggio, 1998; Edelman, 1998a). The computational underpinnings of this alternative approach are discussed elsewhere (section 5.2; see also (Sinha and Poggio, 1996)). For now, we shall take the encroachment of alternative learning models into territory hitherto reserved for structural methods as a license to focus our review on neural, rather than symbolic, computation.

4 Cues for learning

A central question to be addressed in the modeling of a perceptual learning task is that of *supervision*: what are the sources and what is the form of the information that guides the learning process? The usual distinction found in the literature is between supervised models, which require each training stimulus to be accompanied by the desired output, and unsupervised ones, which are able to extract some statistical information from the data, without guidance.

Classifying an experimental setup on the basis of supervision available to the learning model is, however, not always as straightforward as it may seem. Even when the learning process is fully and explicitly controlled by the experimenter, the subject has access to, and is likely to make use of, information that transcends the experimental design. For example, when the subject is required to learn the names of some unfamiliar faces, the ensuing confusion rate will be higher among some

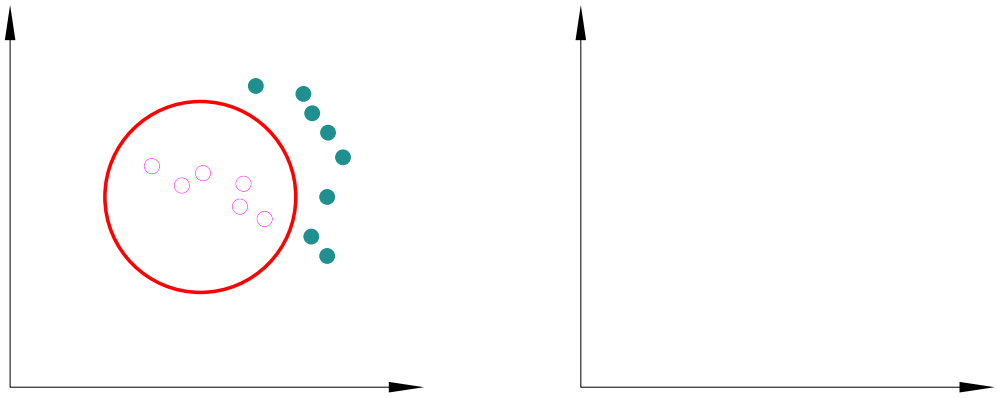
[4] At a certain level of abstraction, the distinction between symbolic and "connectionist" computational paradigms ceases to make sense, as attested by the progress in implementing rules and variables in neural networks; see, e.g., (Ajjanagadde and Shastri, 1991).

faces compared to other.

The reasons for the

4.2 Su ervised learning

Whereas unsupervised algorithms have only the feature-space



task are parameterized by only a few variables (Edelman and Intrator, 1997), and (2) the "front end" of a typical visual system — its *measurement space*[9] — largely preserves the local geometry of the distal problem space (Edelman, 1999).

The first of the two observations that we just stated can be illustrated

with the dimensionality of the embedding space. This includes the Self-Organizing Map algorithm (Kohonen, 1982), and the different varieties of auto-encoders, or bottleneck networks (Cottrell et al., 1987; Leen and Kambhatla, 1994).

The performance of such unsupervised or self-supervised manifold-extracting algorithms can be improved if additional knowledge is brought to bear on the problem. Typically, this is done by making the learning mechanism observe certain invariances known to apply to the problem (Foldiak, 1991; Wiskott, 1998). A particularly simple way to do that is by providing the label of the category to which each stimulus belongs.[11] To see how this information helps the algorithm to isolate the relevant manifold, note that directions orthogonal to it can be effectively specified by forcing stimuli that differ along those directions to be mapped to the same category (Intrator and Edelman, 1997).

5.3 Learning visual category structure (classification)

From the perspective of the task, the main difference between regression and classification is that in the latter, the location of the point within the low-dimensional structure does not matter, while in the former it does. For example, the location of the point representing a face in a face space (the manifold corresponding to the different possible views of the same face) would encode its orientation — a piece of information that should not be discarded. In comparison, in the vernier task, where the problem is that of classification, only the membership in one of the two clusters in the representation space matters to the system.

Despite this difference, the basic considerations identified before in the discussion of regression apply also to classification. In particular, the curse of dimensionality still has to be taken into account. Huber (1985) illustrates this point quantitatively, by showing how difficult it is to find a 3-dimensional Gaussian bump (which could, in terms of Figure 3, right, correspond to one of the class-conditional clusters), when it is embedded in a 10-dimensional space. Although neurally inspired models of learning that are tailored specifically for categorization and mixture estimation do exist (Carpenter and Grossberg, 1990; Carpenter et al., 1992; Williamson,

well-known that this information reveals everything that there is to know about stochastic data, such as the measurements performed by a perceptual system upon the world. The underlying generator of the data (and, therefore, quantities needed for regression or classification) can then be estimated optimally from the density function. However, the first step in this process — the inference of an unconstrained density function from data — is prone to the curse of dimensionality, as shown in the seminal work of Stone (1980; 1982).

In view of this problem, researchers typically take two approaches, which are not mutually exclusive. The first is to make some assumptions about the density function. For example, one may assume that the density function is smooth, then estimate it using splines (Wahba, 1979) or radial basis functions (Poggio and Girosi, 1990). Alternatively, one may assume something about the structure of the density. For example, it may be postulated to belong to an additive model, making it expressible as a sum of functions of some low-dimensional projections of the data (Stone, 1985; Stone, 1986). One may also assume that the density is factorial, namely, a product of marginal densities of one variable (Dayan et al., 1995). The latter two methods do not attempt to reduce the dimensionality of the density function, yet they do make the estimation process more efficient and less prone to the curse of dimensionality.

The second general approach bypasses the problem of density estimation, rather than attempting to solve it. It is based on the observation that for many practical problems only a certain function of the density is required. The hope is that such a function can be easily computed directly from the data, without the need to go via the full density estimation. This happens, e.g., when the desired function is defined over a low-dimensional manifold embedded in the original space, or, more generally, when the desired function has a simpler structure compared to the full density. In such cases, the learning system may attempt to extract the low-dimensional representation of the problem from the data, using an unsupervised approach such as principal component analysis and its generalizations, or using a supervised approach tailored to the desired target function, as it is done in many feed-forward network models.

In all these cases, a model would do well if it applies the methods listed in the preceding sections, which dealt with learning manifold extraction (regression) and clustering (classification). Reverting to those methods means, effectively, that their subsumption under the aegis of density estimation is not practical, unless the estimation algorithm (1) aims for learning a certain target function of the density, which is usually problem-specific, and (2) relies on some prior assumptions about the properties of the desired representation, such as low dimensionality and smoothness (Intrator, 1993).

6 Discussion

The theoretical stance adopted so far in this paper equates learning with the acquisition of efficient representations, a computational procedure that can be regarded as a kind of statistical inference. It may seem that exploring the implications of this stance have lead us away from the gritty details which must be dealt with by any model that aims to simulate human learning behavior. We believe, however, that a good model starts at the top, with a clear notion of what is being modeled, and why. In this concluding section, we recapitulate the links between computational theory and mechanism-level practice in the modeling of perceptual learning, and speculate about the possible future developments in the modeling of perceptual learning.

6.1 On the levels of explanation of learning

The overarching concern in the modeling of a perceptual phenomenon is, of course, getting the performance right. Beyond that, however, there is a considerable variation in what is deemed acceptable: while some comprehensive models treat both the computational (theoretical) and the implementational aspects of the problem, others tend to concentrate on the issues of implementation and mechanism. Models built around neural networks are particularly likely to belong to the second category, going straight from the phenomenology to a hypothesis about the underlying mechanism, perhaps also attempting to emulate along the way the *real* biological neural network.

We illustrate this observation with an example that involves one of the most striking manifestations of perceptual learning, found in the task of detecting a small low-contrast Gabor patch projected onto a certain retinotopically defined location. The detection threshold in this task depends on whether or not the target patch is flanked at a distance by patches of similar orientation and spatial frequency (Polat and Sagi, 1993). Shortly after the effect of the flanking patches has been demonstrated, it turned out to be amenable to learning: the spatial range of the effect (i.e., the maximum effective distance between the target and the flanking patches) grows with practice (see Sagi and Zenger, this volume). Significantly, learning is only possible if the original, untrained range is extended gradually, by exposing the subject to configurations of progressively larger and larger extent (Polat and Sagi, 1994).

A phenomenon such as this seems positively to demand a mechanism-level explanation in terms of receptive fields of retinotopic "units," linked laterally and exerting facilitatory influence on each other; (Polat and Sagi, 1994) offered precisely this explanation for their psychophysical findings. However, as we claimed in the introduction, models formulated primarily in the language of units and connections achieve less than what a model can and should achieve, because they concentrate on the wiring details at the expense of leaving the master plan — the computational goal of the system — out of the picture. To support this argument, let us re-consider the "lateral learning" scenario, keeping in mind the taxonomy of learning paradigms discussed earlier.

Assume for the moment that the goal of the system is to detect the faintest possible line element (a real-life counterpart to a Gabor patch) in a given retinal location. Merely lowering the decision threshold for that location will likely just increase the false-alarm rate there; additional information must be brought to bear on the decision, if it is to be reliable. The presence of other line elements in the vicinity would count as the necessary additional support, if they are compatible with the original hypothesis (i.e., if their orientation is consistent with that of the element whose fate they are about to seal). Thus, the task at hand can be reformulated as that of (literal) *interpolation* between the flanking lines (or extrapolation, if the continuation of an "end-stopped" segment is sought).

The value of this formulation lies in that it brings about the possibility of a uniform treatment for a range of perceptual learning tasks. Indeed, on an abstract level, learning to detect a Gabor patch flanked by similar patterns is now seen to be the same as learning to recognize an object from a novel viewpoint, which is "sandwiched" between two familiar views. The analogy drawn between these two tasks hinges on a parallel between the *view space* of the object on the one hand, and the *"space" space* — that is, the retinal location space — of the Gabor patch on the other hand. Once this analogy is realized, cross-fertilization may occur in both directions. On the one hand, models of object recognition may benefit from postulating a mechanism that carries out interpolation by growing lateral links between neighboring units in a view representation space. On the other hand, models of line detection may benefit from exploring the possibilities originally developed in the context of object recognition (e.g., interpolation with feedforward basis functions).

An edifying perspective on the issue of levels of modeling is provided by recalling some of

the “old-fashioned” models of brain function (and learning) produced by

representation

Carpenter, G. A. and Grossberg, S. (1990). Adaptive resonance theory: Neural network architectures for self-organizing pattern recognition. In Eckmiller, R., Hartmann, G., and Hauske, G., editors, *Parallel Processing in Neural Systems and Computers*, pages 383–389. North-Holland, Amsterdam.

Carpenter, G. A., Grossberg, S., Markuzon, N., Reynolds, J. H., and Rosen, D. B. (1992). Fuzzy ARTMAP: A neural network architecture for incremental supervised learning of analog multidimensional maps. *IEEE Trans. on Neural Networks*, 3:698–713.

Clark, A. (1993). *Sensory qualities*. Clarendon Press, Oxford.

Cottrell, G. W., Munro, P., and Zipser, D. (1987). Learning internal representations from gray-scale images: An example of extensional programming. In *Ninth Annual Conference of the Cognitive Science Society*, pages 462–473, Hillsdale. Erlbaum.

Dayan, P., Hinton, G. E., and Neal, R. M. (1995). The Helmholtz Machine. *Neural Computation*, 7:889–904.

Deutsch, D. (1997). *The fabric of reality*. Allen Lane.

Dretske, F. (1995). *Naturalizing the mind*. MIT Press, Cambridge, MA. The Jean Nicod Lectures.

Duda, R. O. and Hart, P. E. (1973). *Pattern classification and scene analysis*. Wiley, New York.

Duvdevani-Bar, S., Edelman, S., Howell, A. J., and Buxton, H. (1998). A similarity-based method for the generalization of face recognition over pose and expression. In Akamatsu, S. and Mase, K., editors, *Proc. 3rd Intl. Symposium on Face and Gesture Recognition (FG98)*, pages 118–123, Washington, DC. IEEE.

Ebbinghaus, H. (1885). *Memory: A Contribution to Experimental Psychology*. Dover, New York. reprinted 1964; translated by H. A. Ruger and C. E. Bussenius 1913.

Kohonen, T. (1982). Self-organized formation of topologically correct feature maps. *Biological Cybernetics*, 43:59–69.

Kornblith, H. (1985). *Naturalizing epistemology*. MIT Press, Cambridge, MA.

LaBerge, D. (1976). Perceptual learning and attention. In Estes, W. K., editor, *Handbook of learning and cognitive processes*, volume 4, pages 237–273. Erlbaum, Hillsdale, NJ.

Leen, T. K. and Kambhatla, N. (1994). Fast non-linear dimension reduction. In Cowan, J. D., Tesauro, G., and Alspector, J., editors, *Advances in Neural Information Processing Systems*, volume 6, pages 152–159. Morgan Kauffman, San Francisco, CA.

Markson, L. and Bloom, P. (1997). Evidence against a dedicated system for word learning in children. *Nature*, 385:813–815.

Marr, D. (1970). A theory for cerebral neocortex. *Proceedings of the Royal Society of London B*, 176:161–234.

Marr, D. (1971). Simple memory: a theory for archicortex. *Phil. Trans. Royal Soc. London*, 262:23–81.

Marr, D. and Poggio, T. (1977). From understanding computation to understanding neural circuitry. *Neurosciences Res. Prog. Bull.*, 15:470–488.

Minsky, M. and Papert, S. (1969). *Perceptrons*. MIT Press, Cambridge, MA.

Newell, F., Chiroro, P., and Valentine, T. (1999). Recognising unfamiliar faces: The effects of distinctiveness and view. *Q. J. Exp. Psychol.* in press.

Nosofsky, R. M. (1988).

Poggio, T. and Girosi, F. (1990). Regularization algorithms for learning that are equivalent to multilayer networks. *Science*, 247:978–982.

Polat, U. and Sagi, D. (1993). Lateral interactions between spatial channels: suppression and facilitation revealed by lateral masking experiments. *Vision Research*, 33:993–997.

Polat, U. and Sagi, D. (1994). Spatial interactions in human vision: from near to far via experience dependent cascades of connections. *Proceedings of the National Academy of Science*, 91:1206–1209.

Popper, K. R. (1992). *Conjectures and*

Stone, J. V. (1996). Learning perceptually salient visual parameters using spatiotemporal smoothness constraints. *Neural Computation*, 8:1463–1492.

Uttley, A. M. (1959). The design of conditional probability computers. *Information and Control*, 2:1–24.

Valentine, T. (1991). Representation and process in face recognition. In Watt, R., editor, *Vision and visual dysfunction*, volume 14, chapter 9, pages 107–124. Macmillan, London.

Vapnik, V. (1995). *The nature of statistical learning theory*. Springer-Verlag, Berlin.

Wahba, G. (1979). Convergence rates of 'thin plate' smoothing splines when the data are noisy. In Gasser, T. and Rosenblatt, M., editors, *Smoothing Techniques for Curve Estimation*, pages 233–245. Springer Verlag, Berlin.

Walk, R. D. (1978). Perceptual learning. In Carterette, E. C. and Friedman, M. P., editors, *Handbook of Perception*, volume IX, pages 257–298. Academic Press, New York, NY.

Williamson, J. R. (1997). A constructive, incremental-learning network for mixture modeling and classification. *Neural Computation*, 9:1517–1543.

Willshaw, D. J., Buneman, O. P., and Longuet-Higgins, H. C. (1969). Non-holographic associative memory. *Nature*, 222:960–962.

Wilson, M. A. and McNaughton, B. L. (1993). Dynamics of the hippocampal ensemble code for space. *Science*, 261:1055–1058.

Winston, P., editor (1975). *The psychology of computer vision*. McGraw-Hill, New York.

Wiskott, L. (1998). Learning invariance manifolds. In Niklasson, L., Bodén, M., and Ziemke, T., editors, *Proc. Int'l Conf. on Artificial Neural Networks, ICANN'98, Skövde*, Perspectives in Neural Computing, pages 555–560. Springer.