

Epilogue

Neural networks represent a *multidisciplinary* subject with roots in the neurosciences, mathematics, statistics, physics, computer science, and engineering, as evidenced by the diversity of topics covered in this book. Their ability to *learn* from data with or without a teacher has endowed them with a powerful property. This learning property has profound theoretical as well as practical implications. In one form or another, the ability of neural networks to learn from examples (representative of their environment) has made them invaluable tools in such diverse applications as modeling, time series analysis, pattern recognition, signal processing, and control. In particular, neural networks have a great deal to offer when the solution of a problem of interest is made difficult by one or more of the following points:

- Lack of physical/statistical understanding of the problem
- Statistical variations in the observable data
- Nonlinear mechanism responsible for the generation of the data

The new wave of neural networks (since the mid 1980s) came into being because learning could be performed at multiple levels. Neural network based learning algorithms have allowed us to eliminate the need for handcrafted feature extraction in handwriting recognizers. Gradient-based learning algorithms inspired by neural networks have allowed us to simultaneously train feature extractors, classifiers, and contextual processors (hidden Markov models and language models) simultaneously. Because of neural networks we have learning all the way down from pixels to symbols.

Learning pervades every level of intelligent machines in an increasing number of applications. It is therefore befitting that this epilogue concludes the book with final remarks on some intelligent machines and the role of neural networks in building them.

INTELLIGENT MACHINES

With no agreed upon scientific definition of intelligence¹ and due to space limitations, we will not venture into a discussion of what intelligence is. Rather, we will confine our brief exposition to intelligent machines in the context of three specific areas of applica-

tion: pattern classification, control, and signal processing. It is recognized here that there is no "universal" intelligent machine; instead, we have intelligent machines for specific applications.

Much of the research effort on neural networks has focused on pattern classification. Given the practical importance of pattern classification and its rather pervasive nature, and the fact that neural networks are so well suited for the task of pattern classification, this concentration of research effort has been largely the right thing to do. In so doing, we have been able to lay down the foundations of *adaptive pattern classification*. However, we have reached the stage where we think of classification systems in much broader sense if we are to be successful in solving classification problems of more complex and sophisticated nature. Figure 1 depicts the layout of a "hypothetical classification system (Hammerstrom and Rahfuss, 1992). The first level of the system receives sensory data generated by some source of information. The second level extracts a set of features that characterize the sensory data. The third level classifies the features into one or more distinct categories, which are then put into global context by the fourth level. Finally, we may, for example, put the parsed input into some form of a database for an end user. The important features that characterise the system (Fig. 1) are:

- *Recognition*, resulting from the forward flow of information from one level of the system to the next as in a traditional pattern classification system
- *Focusing*, whereby a higher level of the system is able to selectively influence the information processing at a lower level by virtue of knowledge gained from past data

Thus the novelty of the pattern classification system shown in Fig. 1 lies in *knowledge of the target domain* and its exploitation by lower levels of the system to improve overall system performance, given the fundamental constraint of a limited information processing

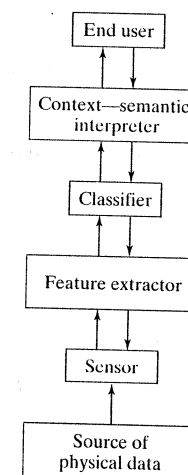


FIGURE 1 Functional architecture of an intelligent machine for pattern classification.

capacity. It is our belief that the evolution of pattern classification using neural networks will be in the direction of creating models that are continually influenced by knowledge of the target domain. We envision a new class of intelligent machines for pattern classification that offer the following attributes:

- Ability to extract *contextual knowledge*, and exploit it through the use of *focusing*
- *Localized* rather than distributed representation of knowledge
- *Sparse* architecture, emphasizing network modularity and hierarchy as principles of neural network design

The realization of such an intelligent machine can only be attained by combining neural networks with other appropriate tools. A useful tool that comes to mind here is the *Viterbi algorithm*, which is a form of dynamic programming designed to deal with *sequential* information processing² that is an inherent characteristic of the system described in Fig. 1. (The dynamic programming algorithm is discussed in Chapter 12.)

Control, another area of application naturally suited for neural networks, is also evolving in the direction of *intelligent control*.³ Autonomy is an important objective of control system designers and intelligent controllers are one way to achieve it. Figure 2 shows a functional architecture for an intelligent autonomous controller with an interface at one end to the process (plant) involving sensing and an interface at the other end to humans and other systems (Antsaklis et al., 1996; Passino, 1996). The system has three functional levels, as summarized here:

1. *Execution level*, which has low-level signal processing and control algorithms for adaptive control and identification.
2. *Coordination level*, which provides the link between the execution and management levels by looking after such matters as tuning, supervision, crisis management, and planning.
3. *Management and organization level*, which provides for the supervision of lower-level functions and management of the interface to the human(s).

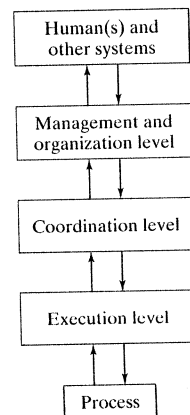


FIGURE 2 Functional architecture of an intelligent machine for control.

Whereas classical control is rooted in the theory of linear differential equations, intelligent control is largely *rule based* because the dependencies involved in its deployment are much too complex to permit an analytical representation. To deal with such dependencies, it is expedient to use the mathematics of fuzzy systems and neural network. The power of *fuzzy systems*⁴ lies in their ability (1) to quantify linguistic inputs, and (2) to quickly give a working approximation of complex and often unknown system input-output rules. The power of neural networks is in their ability to *learn* from data. There is a natural synergy between neural networks and fuzzy systems that makes their *hybridization* a powerful tool for intelligent control and other applications.

Turning next to signal processing, we have yet another fertile area for the application of neural networks by virtue of their nonlinear and adaptive characteristics (Haykin, 1996). Many of the physical phenomena responsible for the generation of *information-bearing signals* encountered in practice (e.g., speech signals, radar signals, sonar signals) are governed by *nonlinear dynamics of a nonstationary and complex nature*, defying an exact mathematical description. To exploit the full information content of such signals at all times, we need intelligent machines for signal processing,⁵ the design of which addresses the following key issues:

- *Nonlinearity*, which makes it possible to extract the higher-order statistics of the input signals
- *Learning and adaptation*, by means of which the system is enabled to learn the underlying physical mechanism of the environment in which it is embedded and to adapt to slow statistical variations in the environment on a continuing basis
- *Attentional mechanism*, whereby, through interaction with the end user or in a self-organized manner, the system is enabled to focus its computing power around a particular point in an image or a particular location in space for more detailed analysis⁶

Figure 3 shows the functional architecture of an intelligent machine for signal processing that involves three levels of operation:

1. *Low-level processing*, the purpose of which is to preprocess the received signal to prepare it for the second level. The preprocessing involves the use of filtering to reduce the effects of noise and other advanced signal processing operations such as *time-frequency analysis*.⁷ The aim of time-frequency analysis is to describe how the spectral content of a signal evolves and to understand what a time varying spectrum is. Specifically, a one-dimensional (temporal) representation of the received signal is transformed into a two-dimensional image with one dimension representing time and the other dimension representing frequency. Time-frequency analysis provides an effective method for highlighting the nonstationary nature of the received signal in a manner far more discernible than in its original temporal form.
2. *Learning and adaptation level*, where memory (of a long-term as well as of a short-term nature) and an attentional mechanism are built into the design of the system. By having the multilayer perceptron, for example, undergo supervised learning with a large enough data set representative of the environment in which the system is embedded, overall statistical information about the environment is

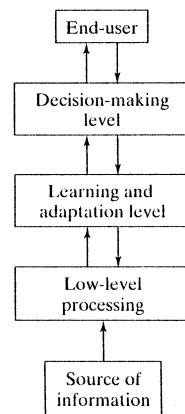


FIGURE 3 Functional architecture of an intelligent machine for signal processing.

stored in the synaptic weights of the network. To account for slow statistical variations in the environment with time, a blind adaptation scheme (i.e., a continuous learning subsystem operating in an unsupervised manner) is appended to the output end of the multilayer perceptron. The learning process also includes provision for an attentional network⁸ whereby the system can focus its attention on important features in the received signal by “gating” the flow of information from lower to higher levels as the need arises.

3. *Decision-making level*, where the final decisions are made by the system. The decision could be whether or not a target of interest is present in the received signal as in radar or sonar, or whether the received bit of information corresponds to symbol 1 or symbol 0 as in digital communications; levels of confidence in the decisions made are also provided.

We do not claim that the systems described here are the only ways in which intelligence could be built into pattern classification, control, and signal processing. Rather, they represent systematic ways in which this important objective could be accomplished. Despite their differences in terms of the domain of application, they do share some common features (Valvanis and Saridis, 1992; Passino, 1996):

- There is a bidirectional flow of information, from lower to higher levels and back.
- Higher levels are often concerned with those aspects of the system’s behavior that are slower in processing time, broader in scope, and longer in horizon time.
- There is increasing intelligence with decreasing precision as we move from lower to higher levels.
- At the higher levels, there is a decrease in *granularity* (i.e., an increase in model abstraction).

We began the discussion of (artificial) neural networks in Chapter 1 by describing the human brain, the source of motivation for neural networks, as a gigantic information processing machine. It is appropriate to conclude the book with a brief exposition of intelligent machines that are the ultimate in information processing by artificial means. The struggle to build intelligent machines continues.

NOTES AND REFERENCES

1. For a philosophical discussion of intelligence from different perspectives, see Ackerman (1990), Albus (1991), and Kosko (1992).
2. The Viterbi algorithm was originally developed by Viterbi (1967) for solving convolutional decoding problems in communication theory. For a tutorial treatment of the Viterbi algorithm, see Forney (1973).
For an application in pattern classification that involves the combined use of the convolutional network (described in Chapter 4) and the Viterbi algorithm, see LeCun et al. (1997, 1998).
3. Intelligent control is discussed in the edited books by White and Sofge (1992), Antsaklis and Passino (1993), Gupta and Sinha (1996), and Tzefestas (1997).
4. Fuzzy set theory was originated by Zadeh (1965, 1973) to provide a mathematical tool for dealing with linguistic variables (i.e., concepts described in natural language). For a treatment of fuzzy logic in book form, see Dubois and Prade (1980). In the book by Kosko (1997), a different viewpoint is taken: Fuzzy systems are viewed as function approximators. Therein it is shown that fuzzy systems can model any continuous function or system provided they use enough rules.
5. A special issue of the 1998 Proceedings of the Institute of Electrical and Electronic Engineers (IEEE) is devoted to the subject of intelligent signal processing (Haykin and Kosko, 1998).
6. A self-organizing system for hierarchical focusing or selective attention is described in Fukushima (1988a). The system is a modification of the layered *neocognitron* also pioneered by Fukushima (1975, 1988b). The system is able to focus attention on an individual character in an image composed of multiple characters or on a greatly deformed character that is also contaminated with noise.
A self-organized attentional mechanism also features in the development of *adaptive resonance theory* (ART) pioneered by Carpenter and Grossberg (1987, 1995). ART for adaptive pattern recognition involves the combination of bottom-up filtering and top-down template matching.
7. For a detailed treatment of the many facets of time-frequency analysis, building on the classical Fourier theory, see the book by Cohen (1995).
For the theory and applications of the Wigner distribution, an important tool for bilinear/quadratic time-frequency representations, see the book by Mecklenbräuker and Hlawatsch (1997).
For a different perspective, where we think in terms of scale in place of frequency, see the book by Vetterli and Koračević (1995) on wavelets and the related topic of sub-band coding.
8. In van de Laar et al. (1997), a neural network model for selective covert visual attention is described. The model is able to learn to focus its attention on important features, depending on the task to be fulfilled, by modulating the flow of information in the preattentive stage.

Bibliography

- Aarts, E., and J. Korst, 1989. *Simulated Annealing and Boltzmann Machines: A Stochastic Approach to Combinatorial Optimization and Neural Computing*, New York: Wiley.
- Abarbanel, H.D.I., 1996. *Analysis of Observed Chaotic Data*, New York: Springer-Verlag.
- Abraham, R.H., and C.D. Shaw, 1992. *Dynamics of the Geometry of Behavior*, Reading, MA: Addison-Wesley.
- Abu-Mostafa, Y.S., 1995. "Hints," *Neural computation*, vol. 7, pp. 639-671.
- Abu-Mostafa, Y.S., 1990. "Learning from hints in Neural Networks," *Journal of Complexity*, vol. 6, pp. 192-198.
- Abu-Mostafa, Y.S., 1989. "The Vapnik-Chervonenkis Dimension: Information Versus Complexity in Learning," *Neural Computation*, vol. 1, pp. 312-317.
- Abu-Mostafa, Y.S., and J.M. St. Jacques, 1985. "Information capacity of the Hopfield model," *IEEE Transactions on Information Theory*, vol. IT-31, pp. 461-464.
- Ackerman, P.L., 1990. "Intelligence." In S.C. Shapiro, ed., *Encyclopedia of Artificial Intelligence*, pp. 431-440, New York: Wiley (Interscience).
- Ackley, D.H., G.E. Hinton, and T.J. Sejnowski, 1985. "A Learning Algorithm for Boltzmann Machines," *Cognitive Science*, vol. 9, pp. 147-169.
- Aiyer, S.V.B., N. Niranjani, and F. Fallside, 1990. "A theoretical investigation into the performance of the Hopfield model," *IEEE Transactions on Neural Networks*, vol. 15, pp. 204-215.
- Aizerman, M.A., E.M. Braverman, and L.I. Rozonoer, 1964a. "Theoretical foundations of the potential function method in pattern recognition learning," *Automation and Remote Control*, vol. 25, pp. 821-837.
- Aizerman, M.A., E.M. Braverman, and L.I. Rozonoer, 1964b. "The probability problem of pattern recognition learning and the method of potential functions," *Automation and Remote Control*, vol. 25, pp. 1175-1193.
- Akaike, H., 1974. "A new look at the statistical model identification," *IEEE Transactions on Automatic Control* vol. AC-19, pp. 716-723.
- Akaike, H., 1970. "Statistical predictor identification," *Annals of the Institute of Statistical Mathematics*, vol. 22, pp. 202-217.
- Albus, J.S., 1991. "Outline for a theory of intelligence," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 21, pp. 473-509.
- Aleksander, I., and H. Morton, 1990. *An Introduction to Neural Computing*, London: Chapman and Hall.
- Allport A., 1989. "Visual attention," In *Foundations of Cognitive Science*, M.I. Posner, ed., pp. 631-682, Cambridge, MA: MIT Press.
- Al-Mashouq, K.A., and I.S. Reed, 1991. "Including hints in training neural nets," *Neural Computation*, vol. 3, pp. 418-427.
- Alspector, J., R.B. Allen, A. Jayakumar, T. Zeppenfeld, and R. Meir, 1991. "Relaxation networks for large supervised learning problems," *Advances in Neural Information Processing Systems*, vol. 3, pp. 1015-1021, San Mateo, CA: Morgan Kaufmann.
- Alspector, J., A. Jayakumar, and S. Luna, 1992. "Experimental evaluation of learning in a neural microsystem," *Advances in Neural Information Processing Systems*, vol. 4, pp. 871-878, San Mateo, CA: Morgan Kaufmann.
- Alspector, J., R. Meir, B. Yuhas, A. Jayakumar, and D. Lippe, 1993. "A parallel gradient descent method learning in analog VLSI neural networks," *Advances in Neural Information Processing Systems*, vol. pp. 836-844, San Mateo, CA: Morgan Kaufmann.
- Amari, S., 1998. "Natural gradient works efficiently in learning," *Neural Computation*, vol. 10, pp. 251-276.
- Amari, S., 1997. Private communication.
- Amari, S., 1993. "A universal theorem on learning curves," *Neural Networks*, vol. 6, pp. 161-166.
- Amari, S., 1990. "Mathematical foundations of neurocomputing," *Proceedings of the IEEE*, vol. 78, pp. 1443-1463.
- Amari, S., 1987. "Differential geometry of a parametric family of invertible systems—Riemannian metric dual affine connections and divergence," *Mathematical Systems Theory*, vol. 20, pp. 53-82.
- Amari, S., 1985. *Differential-Geometrical Methods in Statistics*, New York: Springer-Verlag.
- Amari, S., 1983. "Field theory of self-organizing neural nets," *IEEE Transactions on Systems, Man, and Cybernetics* vol. SMC-13, pp. 741-748.
- Amari, S., 1980. "Topographic organization of nerve fields," *Bulletin of Mathematical Biology*, vol. 42, pp. 339-364.
- Amari, S., 1977a. "Neural theory of association and concept-formation," *Biological Cybernetics*, vol. 22, pp. 175-185.
- Amari, S., 1977b. "Dynamics of pattern formation in lateral-inhibition type neural fields," *Biological Cybernetics*, vol. 27, pp. 77-87.
- Amari, S., 1972. "Characteristics of random nets of analog neuron-like elements," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. SMC-2, pp. 643-657.
- Amari, S., 1967. "A theory of adaptive pattern classifiers," *IEEE Trans. Electronic Computers*, vol. EC-16, pp. 299-307.
- Amari, S., and M.A. Arbib, 1977. "Competition and cooperation in neural nets," in J. Metzler, ed., *System Neuroscience*, pp. 119-165, New York: Academic Press.
- Amari, S., and J.-F. Cardoso, 1997. "Blind source separation—Semiparametric statistical approach," *IEEE Transactions on Signal Processing*, vol. 45, pp. 2692-2700.
- Amari, S., T.-P. Chen, and A. Cichoki, 1997. "Stability analysis of learning algorithms for blind source separation," *Neural Networks*, vol. 10, pp. 1345-1351.
- Amari, S., A. Cichoki, and H.H. Yang, 1996. "A new learning algorithm for blind signal separation," *Advances in Neural Information Processing Systems*, vol. 8, pp. 757-763, Cambridge, MA: MIT Press.
- Amari, S., and K. Maginu, 1988. "Statistical neurodynamics of associative memory," *Neural Networks*, vol. 1, pp. 63-73.
- Amari, S., K. Yoshida, and K.-I. Kanatani, 1977. "A mathematical foundation for statistical neurodynamics," *SIAM Journal of Applied Mathematics*, vol. 33, pp. 95-126.
- Amari, S., N. Murata, K.-R. Müller, M. Finke, and H. Yang, 1996a. "Statistical theory of overtraining—I: cross-validation asymptotically effective?" *Advances in Neural Information Processing Systems*, vol. 8, pp. 176-182, Cambridge, MA: MIT Press.
- Ambros-Ingerson, J., R. Granger, and G. Lynch, 1990. "Simulation of paleo-cortex performs hierarchical clustering," *Science*, vol. 247, pp. 1344-1348.
- Amit, D.J., 1989. *Modeling Brain Function: The World of Attractor Neural Networks*, New York: Cambridge University Press.
- Anastasio T.J., 1995. "Vestibulo-ocular reflex: Performance and plasticity," In M.A. Arbib, ed., *The Handbook of Brain Theory and Neural Networks*, Cambridge, MA: MIT Press.
- Anastasio, T.J., 1993. "Modeling vestibulo-ocular reflex dynamics: From classical analysis to neural networks," in F. Eeckman, ed., *Neural Systems: Analysis and Modeling*, pp. 407-430, Norwell, MA: Kluwer.
- Anastasio, T.J., 1991. "A recurrent neural network model of velocity storage in the vestibulo-ocular reflex," *Advances in Neural Information Processing Systems*, vol. 3, pp. 32-38, San Mateo, CA: Morgan Kaufmann.
- Anderson, J.A., 1995. *Introduction to Neural Networks*, Cambridge, MA: MIT Press.
- Anderson, J.A., 1993. "The BSB model: A simple nonlinear autoassociative neural network," in *Associative Neural Memories* (M. Hassoun, ed.) pp. 77-103, Oxford: Oxford University Press.
- Anderson, J.A., 1988. "General introduction," *Neurocomputing: Foundations of Research* (J.A. Anderson and E. Rosenfeld, eds.), pp. xiii-xxi, Cambridge, MA: MIT Press.
- Anderson, J.A., 1983. "Cognitive and psychological computation with neural models," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. SMC-13, pp. 799-815.