Eqs. (7) and (8) to compute $\bar{\delta}_i^c(k)$ and one term of the summation on the right-hand side of Eq. (9) for case c, and then summing over the cases to obtain the gradient given by Eq. (9).

Once the first-order effects on the terminal cost of weight changes have been determined, a variety of procedures have been proposed in the control and neural-net literature for actually modifying the weights prior to a repetition of the preceding procedure.

The best-known BP results[9] look different from ours but are equivalent. They are stated in terms of quantities $\delta_i^c(k)$ [instead of our $\bar{\delta}_i^c(k)$], and these quantities are interpretable as the negative of the partial derivative of the criterion value with respect to the *net input* to unit i on layer k for case c rather than the positive partial derivative with respect to the output.

Sometimes classification of cases into more than two classes, or the mapping of input vectors into associated output *vectors*, is desired. Then more than one unit is located on the output level for each case. In both instances only a minor modification of the preceding results is required, with the squared-error criterion now involving a summation over all output units for each case.

In a strict mathematical sense, BP has not solved the problem of efficiently determining weights in multilayer networks such that a net will produce stipulated results for a training set of cases. Like any gradient procedure applied to nonlinear and generally nonconvex problems, convergence to local minima (or other stationary points) with nonzero error is possible, even when weights yielding zero error exist. A mathematician would demand a procedure known to yield a solution if one exists [which is the case for the one-layer perceptron using the perceptron learning rule of Eq. (2)] before he or she would consider the problem solved. Additionally, the error function viewed as a function of the weights sometimes has many saddle points that, even if the gradient procedure never exactly reaches one, cause the procedure to take many thousands of steps before convergence on a local minimum. This possibility renders the procedure's efficiency poor and also, thus far, unanalyzable.

Conclusion

Once all of the cases to be learned in a neural-net mapping problem are concatenated into one large network with a vector output, one component for each case, a standard discrete-time optimal-control problem results. The Kelley-Bryson gradient formulas for such problems have been rediscovered by neural-network researchers and termed back propagation. The recursive derivation of these formulas using the chain rule, commonly seen in the neural-network literature, was first used for optimal-control problems by Dreyfus.

References

[1]Dreyfus, H., and Dreyfus, S., "Making a Mind vs Modeling the Brain: Artificial Intelligence Back at a Branchpoint," *Daedalus*, Winter 1988, pp. 15–43.

[2]Dreyfus, H., and Dreyfus, S., *Mind over Machine: The Power of Human Intuition and Expertise in the Era of the Computer*, Free Press, New York, 1988.

[3]Kelley, H. J., "Gradient Theory of Optimal Flight Paths," *ARS Journal*, Vol. 30, No. 10, 1960, pp. 947–954.

[4]Bryson, A. E., and Denham, W., "A Steepest-Ascent Method for Solving Optimum Programming Problems," *Journal of Applied Mechanics*, Vol. 29, No. 2, 1962, pp. 247–257; see also Bryson, A. E., et al., "Determination of the Lift or Drag Program that Minimizes Reentry Heating with Acceleration or Range Constraints Using a Steepest Descent Computation Procedure," Inst. of Aeronautical Sciences, Paper 61-6, 1961.

[5]Bryson, A. E., "A Gradient Method for Optimizing Multi-Stage Allocation Processes," *Proceedings of the Harvard University Symposium on Digital Computers and Their Applications*, April 1961.

[6]Bryson, A. E., and Ho, Y. C., *Applied Optimal Control*, Blaisdell, Waltham, MA, 1969, pp. 43–45.

[7]Dreyfus, S., "The Numerical Solution of Variational Problems," *Mathematical Analysis and Applications*, Vol. 5, No. 1, 1962, pp. 30–45.

[8]Werbos, P., "Beyond Regression: New Tools for Prediction and Analysis in the Behavioral Sciences," Ph.D. Dissertation, Committee on Applied Mathematics, Harvard Univ., Cambridge, MA, 1974.

[9]Rumelhart, D., Hinton, G., and Williams, R., *Parallel Distributed Processing*, Vol. 1, edited by D. Rumelhart, J. McClelland, and the PDP Research Group, MIT Press, Cambridge, MA, 1986, Chap. 8.

Book Announcements

VUKOBRATOVIC, M., BOROVAC, B., SURLA, S., and **STOKIC, D.,** ***Biped Locomotion: Dynamics, Stability, Control and Application,*** Springer-Verlag, Berlin, 1990, 349 pages.

Purpose: This volume is intended for researchers interested in the study of biped gait and its stabilization.

Contents: Dynamics of biped locomotion; synthesis of nonlinear dynamics; control and stability; realizations of anthropomorphic mechanics; appendices.

MEIROVITCH, L., ***Dynamics and Control of Structures,*** Wiley, New York, 1990, 425 pages.

Purpose: Various disciplines involved in the control of structures, namely, analytical mechanics, structural dynamics, and control theory, are represented in this book. The coverage of each discipline is sufficiently detailed to provide a broad picture of the field of structure control.

Contents: Newtonian mechanics; principles of analytical mechanics; concepts from linear systems theory; lumped-parameter structures; control of lumped-parameter systems, classical and modern approaches; distributed-parameter structures, exact and approximate methods; control of distributed structures; review of literature on structural control.

LENT, B., ***Dataflow Architecture for Machine Control,*** Research Studies Press, Somerset, U.K., 1989, 315 pages.

Purpose: This book presents an approach to achieve better price/performance relations for machine control based on the concept of dataflow driven systems.

Contents: Definitions and taxonomy of computer architectures for machine embedded control systems; computer architectures deployed in machine embedded control systems; performance analysis of computer architectures deployed in machine embedded control systems; computer operation in terms of tokens; operational principles for OR tokens; hardware structure supporting the OR dataflow operational principle; OR dataflow operating system; programmability; machine control system based on OR dataflow computer architecture.

SCHIEHLEN, W., (Ed.), ***Multibody Systems Handbook,*** Springer-Verlag, Berlin, 1990, 432 pages.

Purpose: This handbook consists of a collection of programs and software on multibody formalisms contributed by numerous international authors.

Contents: Overview; test examples; description of codes; contributors and distributors.