

Networks and Learning

MIT Industrial Liaison Program

Barry Kort

■ On 15–16 November 1989, I attended the Massachusetts Institute of Technology (MIT) Industrial Liaison Program entitled “Networks and Learning.” The topic was neural networks, their power, potential, and promise. A dozen distinguished professors and researchers presented informative and entertaining talks to an audience of technically minded business executives and industrial researchers who subscribe to MIT’s popular series of symposia offered through their Industrial Liaison Program. This informal report encapsulates the two-day event with a brief summary of each talk.

The Massachusetts Institute of Technology (MIT) Industrial Liaison Program, “Networks and Learning,” was held 15–16 November 1989 at MIT. A dozen distinguished professors and researchers presented informative and entertaining talks to an audience of technically minded business executives and industrial researchers who subscribe to MIT’s popular series of symposia offered through its Industrial Liaison Program. This workshop report contains a brief summary of each talk.

Thomaso Poggio of the MIT Department of Brain and Cognitive Sciences opened the symposium by reviewing the history of advances in the field. About every 20 years, there is an epidemic of activity lasting about 12 years, followed by about 8 years of inactivity. Sixty years ago, the Gestalt school began in Europe. Forty years ago, cybernetics emerged in the United States. Twenty years ago, perceptrons generated a flurry of research. Today, *neural networks* (highly interconnected structures of relatively simple units with algebraic connection weights) represent the latest breakthrough in this series.

Leon Cooper, co-director of the Center for Neural Science at Brown University, spoke on neural networks in real-world applications. Neural nets learn from examples. Give them lots of examples of input-output pairs, and they build a smooth map-

ping from the input space to the output space. Neural nets work best when the rules are vague or unknown. The classical three-stage neural net makes a good classifier. It can divide the input space into arbitrarily shaped regions. At first, the network just divides the space into halves and quarters, using straight-line boundaries (*hyperplanes* for the mathematically minded). Eventually (and with considerable training), the network can form arbitrarily curved boundaries to achieve arbitrarily general classification. Given enough of the critical features on which to reach a decision, networks have been able to recognize and categorize diseased hearts from heartbeat patterns. With a sufficiently rich supply of clues, the accuracy of such classifiers can approach 100 percent. Accuracy depends on the sample length of the heartbeat pattern—a hurried decision is an error-prone decision.

Ron Rivest, associate director of MIT’s Laboratory for Computer Science, surveyed the theoretical aspects of learning and networks. He addressed the question, “How do we discover good methods of solution for the problems we wish to solve?” In studying neural networks, he noted their strengths and characteristics: the ability to learn from example, expressiveness, computational complexity, sample-space complexity, and the ability to learn a mapping. The fundamental unit of a neural network is a linear adder, followed by a threshold trigger. If the algebraic sum of the input signals exceeds threshold, the output signal fires. Neural nets need not be constrained to Boolean signals (zero/one) but can handle continuous analog signal levels. In addition, the threshold trigger can be relaxed to an S-shaped response. Rivest said that any continuous function mapping the unit interval $[-1, 1]$ into itself can be approximated arbitrarily well with a

three-stage neural network. (The theorem extends to the Cartesian product: The mapping can be from an m -fold unit hypercube into an n -fold unit hypercube.) Training the neural net amounts to finding the coefficients that minimize the error between the examples and the neural network’s approximation. The so-called error back-propagation algorithm is mathematically equivalent to least squares curve fitting using steepest descent. Although this method works, it can be slow. In fact, training a three-stage neural network is an NP-complete problem—the work increases exponentially with the size of the network. The classical solution to this dilemma is to decompose the problem into smaller problems (subproblems), each solvable by a smaller system. Open issues in neural network technology include the incorporation of prior domain knowledge and the inapplicability of powerful learning methods such as Socratic-style guided discovery and experimentation. There is a need to

Today, neural networks . . . represent the latest breakthrough . . .

merge the statistical paradigm of neural networks with the more traditional knowledge representation techniques of the analytic and symbolic approaches.

Terry Sejnowski, director of the Computational Neurobiology Laboratory at the Salk Institute for Biological Studies, gave a captivating lecture entitled “Learning Algorithms in the Brain.” Sejnowski, who studies biological neural networks, has witnessed the successful reverse engineering of several complete systems. The vestibular ocular reflex is the feed-forward circuit from the semicircular canals of the inner ear to the eye muscles that allow us to fix on a target even as we move and bob our heads. (If you shake your head as you read this sentence, your eyes can remain fixed on the text.) This circuit has been around for hundreds of millions of years, going back to our reptilian ancestors. It is found in the brain stem and operates with only a

7-millisecond delay. (Tracking a moving target is more complex, requiring a feedback circuit that taps into the higher cognitive centers.) The vestibular ocular reflex appears to be overdesigned, generating opposing signals that at first appear to serve no function. Only recently, a veteran researcher finally explained how the dynamic tension between opposing signals allows the long-term adaptation to growth of the body and other factors (such as new eyeglasses) that could otherwise defeat the performance of the reflex.

Sejnowski also described the operation of one of the simplest neurons, found in the hippocampus, that mediates long-term memory. The Hebb's synapse undergoes a physiological change when the neuron fires during a simultaneous occurrence of stimuli, representing the input-output pair of a training sample. After the physiological change, the neuron becomes permanently sensitized to the input stimulus. The Hebb's synapse seems to be the foundation for superstitious learning.

After a refreshing lunch of cold roast beef and warm conversation, Thomaso Poggio returned to the

podium to speak on networks for learning: a vision application. He began by reviewing the theoretical result that equates the operation of a two-layer neural network to linear regression. To achieve polynomial regression, one needs a three-layer neural network. Such a neural net can reconstruct a (smooth) hypersurface from sparse data. (An example of a nonsmooth map would be a telephone directory that maps names into numbers. No smooth interpolation will enable you to estimate the telephone number of someone whose name is not in the directory.) Poggio explored the deep connection between classical curve fitting and three-stage neural networks. The architecture of the neural net corresponds to the so-called hyperbasis functions, which are fitted to the training data. A particularly simple but convenient basis is a Gaussian function centered on each sample x value. The interpolated y value then is just the average of all the sample y values weighted by their Gaussian multipliers. Thus, the nearest neighbors to x are averaged to estimate the output, $y(x)$. For smooth maps, such a scheme works well.

Richard Lippmann of the MIT Lincoln Laboratory spoke on neural network pattern classifiers for speech recognition. Historically, classification has progressed through four stages—probabilistic classifiers using linear discriminant functions, hyperplane separation using piecewise linear boundaries, receptive field classification using radial basis functions, and the new exemplar method using multilayer perceptrons and feature maps. Surveying and comparing alternate architectures and algorithms for speech recognition, Lippmann reviewed the diversity of techniques, comparing the results, accuracy, speed, and computational resources required. From the best to the worst, they can differ by orders of magnitude in cost and performance.

Michael Jordan of MIT's Department of Brain and Cognitive Science spoke on adaptive networks for motor control and robotics. Much progress has been made in this field over the last five years, but neural nets do not represent a revolutionary breakthrough. The inverse problem in control theory is classical: Find the control sequence that will drive the system from the current state to the

goal state. It is well known from cybernetics that the controller must compute (directly or recursively) an inverse model of the forward system. This problem is equivalent to diagnosing cause from effect. The classical solution is to build a model of the forward system and let the controller learn the inverse through unsupervised learning (playing with the model). The learning proceeds incrementally, corresponding to back propagation or gradient descent based on the transposed Jacobian (first derivative). This approach is essentially how humans learn to fly and drive using simulators.

Danny Hillis, founding scientist of Thinking Machines Corporation, captured the audience with a spell-binding talk entitled "Intelligence as an Emergent Phenomenon." Hillis began with a survey of computational problems well suited to massively parallel architectures—matrix algebra and parallel search. He used the biological metaphor of evolution as his model for massively parallel computation and search. Because the evolution of intelligence is not studied as much as the engineering approach (divide and conquer) or the biological approach (reverse engineer nature's best ideas), Hillis chose to apply his connection machine to the exploration of evolutionary processes. He invented a mathematical organism (a *ramp*) that seeks to evolve and perfect itself. A population cloud of these ramps inhabits his connection machine, mutating, evolving, and competing for survival of the fittest. Hillis's color videos showed the evolution of the species under different circumstances. He found that the steady state did not generally lead to a 100-percent population of perfect ramps; rather, two or more immiscible populations of sub-optimal ramps formed pockets with seething boundaries.

Hillis then introduced a species of parasites that attacked ramps at their weakest points so that stable populations would eventually succumb to a destructive epidemic. The parasites did not clear the way for the emergence of perfect and immune ramps; rather, the populations cycled through a roiling rise and fall of sub-optimal ramps, still sequestered in camps of Gog and Magog. The eerie resemblance to modern geopolitics and classical mythology was palpable and profound.

John Wyatt of the MIT Department

of Electrical Engineering and Computer Science closed the first day's program with a talk entitled "Analog VLSI Hardware for Early Vision: Parallel Distributed Computation without Learning." Wyatt's students are building analog devices that can be stimulated by focusing a scene image onto the surface of a chip. His devices for image processing use low-precision (about eight bits) analog processing based on the inherent bulk properties of silicon. His goal is to produce chips that cost \$4.95. One such chip can find the fixed point when the scene is zoomed. (Say you are approaching the back of a slow moving truck. As the back of the truck looms larger in your field of view, the fixed point in the scene corresponds to the point of impact if you fail to slow down.) Identification of the coordinates of the fixed point and the estimated time to impact are the output of this chip. Charged-coupled devices and other technologies are being transformed into image-processing devices that can perform such tasks as stereo depth estimation, image smoothing and segmentation, and motion vision.

The second day of the symposium focused on the Japanese, European, and American perspectives for the development and application of neural nets. Shun-Ichi Amari of the Department of Mathematical Engineering and Information Physics at the University of Tokyo explored the mathematical theory of neural nets. Whereas conventional computers operate on symbols using programmed sequential logic, neural nets correspond to intuitive styles of information processing—pattern recognition, dynamic parallel processing, and learning. Amari explored neural network operation in terms of mathematical mapping theory and fixed points. Here, the fixed points represent the set of weights corresponding to the stable state after extensive training.

Wolfram Büttner of Siemens Corporate Research and Development discussed several initiatives in Europe to develop early commercial applications of neural net technology. Workpiece recognition in the robotic factory and classification of stimuli into categories are recurring themes here. There is also interest in unsupervised learning (playing with models or exploring complex environments), decision support systems (modeling, prediction, diagnosis, sce-

nario analysis, optimal decision making with imperfect information), and computer languages for neural network architectures. Büttner described NeuroPascal, an extension to Pascal for parallel neurocomputing architectures.

Scott Kirkpatrick, manager of workstation design at IBM's Thomas J. Watson Research Center, explored numerous potential applications of neural nets as information-processing elements. These neural nets can be used as filters, transformers, classifiers, and predictors. Commercial applications include routine processing of high-volume data streams, such as in credit checking and programmed arbitrage trading.

Neural nets are also well suited to adaptive equalization, echo cancellation, and other signal-processing tasks. Science Applications International Corp. is using them in its automated luggage inspection system to recognize the telltale signs of suspect contents of checked luggage. Neurogammon 1.0, which took two years to build, plays a mean game of backgammon, beating all other machines and giving world class human players a run for their money. Hard problems for neural nets include three-dimensional object recognition in complex scenes, natural language understanding, and *database mining* (theory construction). Today's commercially viable applications of neural nets could only support about 200 people. It will be many years before neurocomputing becomes a profitable industry.

Marvin Minsky, MIT's Donner Professor of Science, gave an entertaining talk on future models. The human brain has over 400 specialized architectures and is equivalent in capacity to about 200 Connection Machines™ (Model CM-2). There are about 2000 data buses interconnecting the various departments of the brain. As one moves up the information-processing hierarchy, one begins at sensory motor and advances through concrete thinking, operational thinking, and other stages and arrives at formal thinking, the highest cognitive stage. A human subject matter expert who is a world-class master in his field has about 20,000 to 50,000 discrete chunks of knowledge. Among the computational paradigms found in the brain are space frames (for visual information), script frames (for stories), transframes (for mapping between frames), K-

JA ad

Carnegie Group ad

lines (recurring themes in a person's lifetime experiential learning curve), semantic networks (for vocabulary and ideas), trees (for hierarchical and taxonomic knowledge), and rule-based systems (for bureaucrats). Minsky's theory is summarized in his latest book, *Society of Mind* (Simon & Schuster, 1986). Results with neural networks solving interesting problems, such as backgammon moves or freshman calculus problems, reveal that we don't always know which problems are hard. It appears that a problem is hard until somebody discovers an easy way to solve it. At this point, it's deemed trivial. Referring to intelligence, Minsky says that humans are good at what humans do: "A frog is very good at catching flies. And you're not."

The afternoon panel discussion, led by Patrick Winston, provided the speakers and audience another chance to visit and revisit topics of interest. That commercial neural networks are not solving profoundly deep and important problems was a source of dismay to some, who thought that we had enough programmed trading and credit checking going on already, and we don't need

more robots turning down our loans and sending the stock market into instability. The deeper significance of the symposium was that research in neural networks is stimulating the field of brain and cognitive science and giving us new insights into who we are, how we came to be this way, and where we can go if we use our higher cognitive functions to our best advantage.



Barry Kort is a visiting scientist at BBN Laboratories in Cambridge, Massachusetts, where he is engaged in educational technology and research, specializing in cognitive science, scientific visualization, and computer-animated simulation modeling. He holds an M.S. in electrical engineering and a Ph.D., both from Stanford University, and is a member of the American Association for Artificial Intelligence and the Institute of Electrical and Electronic Engineers. In addition, he works as a volunteer in the Discovery Room at the Boston Museum of Science.